DEVELOPMENT OF A PRACTICAL MODEL FOR SCHOOL LEADERS USING ELEMENTARY STUDENT DATA TO PREDICT HIGH SCHOOL DROPOUT RISK

A Dissertation presented to the Faculty of the Graduate School
University of Missouri

In Partial Fulfillment
Of the Requirements for the Degree

Doctor of Education

---

by
NATHAN HOVEN

Dr. Cindy MacGregor, Dissertation Supervisor

MAY 2013

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

DEVELOPMENT OF A PRACTICAL MODEL FOR SCHOOL LEADERS USING ELEMENTARY SCHOOL DATA TO PREDICT HIGH SCHOOL DROPOUT RISK

presented by Nathan D. Hoven,

a candidate for the degree of doctor of education,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Cynthia MacGregor

Dr. Robert Watson

Dr. Robert Hornberger

Dr. Kim Finch

DEDICATION

This dissertation is dedicated to my wonderful wife Kelly. Your love and support throughout the EdD program and my work on this dissertation was invaluable. You will never fully know how much I needed and appreciated your help. The long nights - at class, traveling, or sitting in front of a computer – and you took care of everything at home so I could take care of school. With three young kids at home this was no easy task, and you handled it all with a smile on your face. I love you so much, and this dissertation (in fact, my whole degree) is yours as well as mine.

Speaking of the kids, I owe you my gratitude as well. For the past four years you have had to share me with “school.” For all the nights and weekends I couldn’t spend the time with you that you (or I) wanted, and still you were always ready for me as soon as I was available. Nicholas, Zachary, and Danielle – I love you all and I thank you for helping to keep me grounded in what’s truly important. Nicholas, I love how you challenge me with your quick wit. Zachary, I love how you make me laugh. Danielle, I love how you know just when I need a hug – and don’t worry, I’ll keep carrying you until I just can’t pick you up anymore.

Finally, I want to recognize my late mom and dad, Dean and Susan Hoven. You raised me to be curious, independent, and motivated – and all of that paid off so much through these last few years. You both died far too soon (mom just this semester), and I really wish you could’ve been here to see me finish this journey. I always tried to make you proud, and I think you would’ve enjoyed this. I miss and love you both so much.

ACKNOWLEDGEMENTS

The past four years have been an incredible journey of personal growth and discovery for me. From the first summer spent in Columbia to this final product, I have worked so hard and learned so much. While the classes and papers required a lot of time sitting alone reading journals or writing at my computer, there are so many people that played significant roles in my success.

First, I want to thank Dr. Cindy MacGregor – my teacher and advisor. Dr. Mac always spoke just the right language for me. Whether she was praising me for work I did well, encouraging me to improve work I didn't, or just generally talking me through the program – she was always the support I needed. I am proud of this final accomplishment – this dissertation – and her influence is all over it. I hope I've done her proud.

And to Dr. Robert Watson, my other teacher and mentor. Dr. Watson had the background and practical experience that aligned so well with my career. He always asked the right questions to provoke my thoughts, or just to let me know he was truly interested in how things were going for me. I enjoyed giving him my weekly updates at class and hearing his wisdom. For this dissertation Dr. Watson continued to offer me the practical guidance I needed to improve the final product, just as I have come to appreciate from him.

To my additional committee members, Dr. Kim Finch and Dr. Robert Hornberger – your help has been appreciated so much as well. From the technical corrections (data were!) to the fresh perspectives only someone separated from K-12 education can provide – you offered such great insights to make this dissertation better. The final result is so much stronger because of your influence.

I had another committee that needs to be thanked, too. Through many hours of class and homework, my satellite group in Lebanon – Randy Caffey, Jodi Elder, and Jenny Flatt – was always there for me. I always enjoyed our conversations and how we all brought out the best in each other. Even though you all finished before me, I'm proud that we traveled this road together, and I'm happy to finally join you.

I also need to acknowledge the support of my administration from the Rolla Public Schools. It was Dr. Giger who first encouraged me to pursue this program and gave me his blessing to spend so much time away from school. After Dr. Giger's retirement, Dr. Zalis didn't hesitate to offer me the same support and encouragement. They never wavered in their support, even as I had to miss time at work on many occasions.

And finally to my newest administrators in the Wentzville School District. They have been supportive of me as I have had to miss work to wrap up the final stages of my program. I am proud to complete this dissertation while being part of such a great school district.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Many researchers have identified the myriad of concerns that frequently affect people who drop out of school prior to high school graduation. These include increased risks of lower income, need for welfare support, unemployment, and criminal activity (Alexander, Entwisle, & Horsey, 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh, Suh, & Houston, 2007). School leaders have a keen interest in helping all students successfully complete school, thereby reducing the risk of these issues occurring later in life.

In an effort to help students avoid these potential risks, school leaders have tried to identify students at risk of dropping out of school so they can intervene and help the students persist to graduation. Efforts to identify potential dropouts generally began at the high school level, but Bowers (2010) noted interventions used that late in a student's education are rarely effective. Suh et al. (2007) discussed how interventions are more effective when put into place early in students' educational careers, while Entwisle and Alexander (1993) argued that most students start elementary school with a clean slate before their pattern of performance tends to lead them toward eventual success or failure at completing high school.

This quantitative study examined a sample of 222 students who entered high school in the same cohort in an effort to identify predictors of high school dropouts. The study attempted to develop a practical model built from elementary school data that would predict the number of high school credits earned for each student, thereby giving school leaders a measure they could analyze to identify which students were at risk of eventually dropping out. Through this process, the study aimed to provide school leaders

with a tool to identify potential dropouts during elementary school, giving them a better chance of applying interventions that were more likely to be successful than those applied later in students' educational careers.

This study also focused on data easily recovered from typical school records. The intention was to build a practical model that could easily be developed from school data and applied to all students in a grade level. The variables considered for this study were attendance, core grades, discipline events, standardized test scores, socioeconomic status, grade retention, and reading level.

The findings of the study were that the earlier the model was developed, the less accurately it predicted high school credits earned. This was consistent with prior research noting that earlier identification efforts tend to be less reliable than later efforts (Bowers, 2010). In addition, the models developed in this study were not very accurate for any of the grade levels considered, from fourth through eighth grades. One positive outcome of the prediction models was the relative success with which they predicted dropouts for students predicted to earn extremely low numbers of credits.

Further study can be conducted on more complete data sets to determine if these models could be improved. In addition, for any attempts to identify students at risk of dropping out of school, appropriate interventions must be developed if school leaders are to try to keep these students in school through high school graduation. This study has offered some insights to aid in this further research.

CHAPTER ONE

INTRODUCTION TO THE STUDY

Many researchers have identified the myriad of concerns that frequently affect people who drop out of school prior to high school graduation. These include increased risks of lower income, need for welfare support, unemployment, and criminal activity (Alexander, Entwisle, & Horsey, 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh, Suh, & Houston, 2007). School leaders have a keen interest in helping all students successfully complete school, thereby reducing the risk of these issues occurring later in life.

One of the main strategies school personnel have used to prevent dropouts is to develop academic interventions targeted at students identified as at risk of dropping out of school. Christenson and Thurlow (2004) suggested interventions should be targeted to the specific deficiencies students experience. They found these kinds of specified interventions can show some success at preventing dropouts. Bowers (2010) noted, however, that interventions used in a high school setting are rarely effective. He pointed out that often the issues that lead to dropping out are too entrenched in high school students for interventions to reverse the course toward dropping out of school. Gleason and Dynarski (2002) drew attention to another aspect of interventions. They found that many intervention efforts were aimed at the wrong students because the identification efforts either falsely predicted future graduates as dropouts or failed to identify students who eventually did drop out of school. If potential dropouts are not correctly identified, efforts to intervene will be unsuccessful.

These types of issues have led school leaders to look for ways to more successfully identify potential dropouts and how to more successfully intervene once those students are identified. Entwisle and Alexander (1993) discussed the importance of early experiences in students' educations. They argued that while students may all start elementary school with a clean slate, they quickly build a pattern of performance that may lead to eventual success or failure at completing high school (Alexander et al., 1997). Preventing dropouts has been shown to be more successful when targeted interventions are in place early in students' educational careers. Early identification of at-risk students leads to more effective intervention and prevention (Suh et al., 2007). Bowers (2010) noted, while middle school-based interventions are much more successful than high school-based interventions, middle school predictors are less reliable. This puts school leaders in a difficult position where they are trying to help these students earlier but cannot reliably identify them.

Efforts to reduce the number of high school dropouts have therefore started to focus on accurate earlier identification (Alexander et al., 1997; Battin-Pearson, Newcomb, Abbott, Hill, Catalano, & Hawkins, 2000; Bowers, 2010; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). If reliable predictors can be used for earlier identification of students at risk of dropping out, intervention and prevention efforts will have more time to be effective. Interventions may have a chance to be effective before destructive educational patterns become entrenched for these students.

Statement of the Problem

As stated earlier, many problems occur at higher rates for students who drop out of school prior to high school graduation compared to those who graduate from high school. These include higher incidents of adults earning lower income, needing welfare support, struggling to maintain employment, and engaging in criminal activity (Alexander et al., 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). School leaders have a deep interest in helping to avoid these potential problems by reducing the number of students who drop out before finishing high school.

Many models have been developed through research to identify potential dropouts. While most of the models offer positive and practical results, they are all lacking in some way as well. Some of them are lacking because the model did not create early enough predictions to maximize effectiveness (Fitzpatrick & Yoels, 1992; Reschly & Christenson, 2006). Others are lacking because the research revealed the model did not predict dropouts accurately enough to be useful (Janosz, LeBlanc, Boulerice, & Tremblay, 1997; Zvoch, 2006) or the results were too narrow to be practical (Annie E. Casey Foundation, 2010; Caraway, Tucker, Reinke, & Hall, 2003). Some models are skewed by other variables the researchers were not able to take into account (Battin-Pearson et al., 2000; Suh et al., 2007), and others utilized a labor-intensive process that would be impractical for schools to apply to all students (Balfanz, Herzog, & Mac Iver, 2007). Each of these models highlights the need for a better model, yet offers useful information to build upon in pursuit of that model.

This leads to the main problem addressed by this study. While educators feel compelled to address the school dropout problem, there are currently no practical tools to reliably and accurately identify potential dropouts at an early enough age to successfully intervene. School personnel have a need for such a tool so they can begin developing and implementing intervention strategies to help these students stay in school through graduation.

Purpose of the Study

The purpose of this quantitative study was to develop a model for educators to use to identify potential future dropouts as early as fourth grade. Fourth grade was chosen for two main reasons. First, a main goal of the study was to develop a prediction model that could be used as early as elementary school. Second, in working with school administrators to identify a district with available data in the archives, fourth grade was the earliest any of the districts had available data. The method of identifying potential dropouts was to predict the number of high school credits earned. The district in the study required 24 credits to graduate from high school, so students predicted to earn less than 24 credits could be considered to be at risk of dropping out of school. Successful development of this model would allow school leaders to implement earlier interventions for identified students, ideally preventing these students from dropping out of school. For this model to be useful for school personnel, it must be accurate and practical.

Existing prediction models have often utilized personal surveys completed by counselors, teachers, or parents, leading to instruments that become cumbersome, impractical, and subjective. It then becomes difficult for schools to routinely use the models to identify students in need of intervention (Balfanz et al., 2007). The model developed in this study was designed to use school-report generated data as input variables to predict eventual numbers of high school credits earned, giving schools an efficient and practical method of identifying students at risk of dropping out in the future.

Research Questions

For this quantitative study three questions will be used to guide the research:

1. What combinations of school and family variables, such as attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention, are the best predictors of the number of high school credits earned?
2. How early are these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) reliable predictors of the number of high school credits earned?
3. How accurately and reliably will a model developed from these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) predict the number of high school credits earned if used as early as fourth grade?

Conceptual Underpinnings for the Study

The conceptual framework for this study is built around how school interventions can affect whether or not students drop out of school before graduating from high school. This brings two factors into the discussion. First, students tend to enjoy school as early elementary students (Alexander et al., 1997), and, unfortunately for some, over time various factors can change that. Second, school leaders have direct control over some, but not all, of the factors that lead students to drop out of school (Stearns & Glennie, 2006). Both of these factors are supported by theory.

The conceptual framework guiding this study is based on two main theories. The first is Erikson’s *Stages of Psychosocial Development* (Franks, 2007). While Erikson described multiple stages of development, one specific stage contributed to the development of the conceptual framework: *Industry vs. Inferiority*. Erikson identified this as a stage that generally occurs for children during elementary to early middle school years.

An important aspect of this stage is the fact that children are of school age when they enter it, meaning children generally enter this stage after having started their school careers (Franks, 2007). During this time, children are learning to practice independence and ideally gaining self-confidence. As Industry vs. Inferiority applies to a school setting, children are learning how to please their teachers through their efforts at producing good work or behaving appropriately. The industriousness children begin to display leads to positive recognition from their teachers, parents, and other adults. This recognition helps the child gain self-confidence, leading to future positive activities.

For some children, unfortunately, their efforts do not lead to positive recognition. Some may be discouraged or mocked when they are unsuccessful, while others may just find it difficult to please their teachers, parents, or other adults. This can quickly lead to a feeling of inferiority. Feelings of inferiority can then lead to a lack of motivation or effort. In a school setting, this can theoretically lead a student to become increasingly disconnected from school, and this may be a path from which the student cannot recover, ultimately leading to the student dropping out of school.

The concept of students developing these feelings of inferiority was essential to guiding the current study. Because many interventions have typically been aimed at older students, often the negative characteristics some students experience are too engrained for interventions to overcome them (Bowers, 2010). The current study aimed to identify and intervene earlier for these students, ideally affecting them before feelings of inferiority potentially surface in the first place.

The second theory that contributed to the development of the framework was the notion of pull-out factors and push-out factors that lead to students dropping out of school (Stearns & Glennie, 2006). This theory separates factors that can lead students to drop out into two main categories: outside influences that lead a student to drop out in pursuit of perceived benefits (i.e., pull-out factors) and school-based factors that may drive students to remove themselves from the school setting (i.e., push-out factors).

Pull-out Theory

Pull-out theory suggests students might be influenced to leave school based on factors like the job market, family responsibilities, or peer relationships (Stearns & Glennie, 2006). The theory also notes that pull-out factors tend to be logic-based decisions for students. For example, a student may feel the need to work to support his or her family, or the wages from a job may be perceived to be of greater benefit than completing school. Generally, pull-out factors do not arise from situations schools can affect.

While pull-out factors can certainly contribute to students' decisions to drop out of school, these factors were not considered for this study due to the nature of the model being developed. The model for this study was built by studying school-reported variables, which were more generally described as push-out factors. Additionally, because the model was developed to be practical for educators to apply to entire populations of students, it was necessary to use data that were easy to collect. Pull-out factors generally do not fit this description.

Push-out Theory

Push-out theory, in contrast, focuses on factors controlled by the school. This can include disciplinary consequences, grading policies, teacher/student interactions, or building climate. This theory notes that these kinds of structures can lead some students to view school negatively, thereby increasing the likelihood the students may seek to remove themselves from the setting (Stearns & Glennie, 2006). It is important to point out that some of these factors may be consciously developed structures within the school, capable of being altered if desired.

Push-out theory has largely driven the choice of predictors for the current study as well. While they are factors school leaders can often control, they are also routinely tracked by educators, helping to create consistent data from which to guide decisions. In contrast, pull-out factors are frequently situations school leaders do not track and may know nothing about (Stearns & Glennie, 2006).

The conceptual framework guiding the research in this study was developed from these two main theories for two reasons. The first reason is the research is focused on early predictors of dropouts in an effort to identify at-risk students before feelings of inferiority might lead to thoughts of dropping out of school. Early identification can lead to early intervention, thereby helping students achieve success before negative school views become entrenched. The second reason is the research is focused on school-based factors that lead to students dropping out because schools can have a degree of control over those factors and how they affect student perceptions of school.

Limitations and Assumptions

This study used longitudinal artifact data to examine characteristics of students from a single graduation cohort, but the data being examined were from their years as elementary students. Examining student data that span 10 or more years creates some limitations and requires some assumptions.

Limitations

A major limitation of this study was the selection of a sample that included only students who had been in the same district throughout the years accounted for in the data collection. It can be very difficult or even impossible to track students who changed districts at any point from elementary school through graduation or when they dropped out of school. While this necessitated a constraint on the sample selection, it could have potentially affected the results of the study due to one major issue: student mobility itself is often identified as a predictor of future dropouts (Jimerson, Anderson, & Whipple, 2002; Montes & Lehmann, 2004; Rumberger & Ah Lim, 2008; Suh et al., 2007). As a result, mobility was not considered as a predictor for the model developed in this study even though it is often shown to be a predictor of future dropouts.

Another limitation of this study was whether or not the results could be generalized to other districts. The study was performed in a suburban Midwest district of roughly 5500 students. The district was made up 57% white students, 32% black students, 7% Hispanic students, and small percentages of various ethnicities. In the district, 43% of students qualified for free or reduced lunch. The results of the study can not necessarily be generalized to other districts that differ demographically from the district in this study. This will limit the usefulness of the study for districts with different geographic, demographic, or socioeconomic compositions. For those districts with similar make-ups, however, this study should provide information to help identify potential dropouts as early as elementary school so schools can begin providing interventions to prevent these students from ultimately dropping out of school.

Assumptions

One major assumption of this study is that enough students in the population attended the same school district since elementary school so the sample size for the study would be of sufficient size. Suburban mobility rates tend to hover around 15% per year for students, with the number reaching anywhere from 30%-40% of students moving two or more times by eighth grade (Miami-Dade County Public Schools, 2007; Rumberger & Larson, 1998). If these estimates are moved even higher, to 50% or 60%, that still leaves a sample size of 160-200 students, more than enough according to Field (2009), who estimated a regression sample would need 10-15 subjects per variable used. With seven independent variables, this would give a minimum sample size of 70-105 subjects. The sample size would still satisfy Field's estimates even if the mobility rate topped 70%, making the assumption of a sufficient sample size reasonable.

Another assumption in this study was that the definition of the term dropout would yield valid results. As stated in the next section, the term dropout is defined differently in almost every study. With so many competing definitions for the term, it is difficult to predict the subtle effects that each definition could have on data analysis. The definition used in this study is grounded in research, and there is rationale for defining it this way.

A final assumption is that data collected from ten or more years ago were accurate. Care was taken to select a school district that had used the same student information system for the duration of the years covered in the study in an effort to minimize potential inaccuracies. Given this choice of district, it was assumed the data would be accurate.

Definition of Key Terms

As with any study, there are terms that will be used in this study that have specific meanings. While they may be used differently in other studies or contexts, these terms will be given fixed definitions specific to this study.

At-risk. The term at-risk was used in a general sense in this study to identify students who exhibited some characteristics that could have led to dropping out of school before high school graduation (Suh et al., 2007).

Attendance. Attendance was used as a percentage of possible school time attended during a given school year. For each student, the amount of time attended was divided by the amount of total time school was in session for that year to develop an attendance percentage (Suh et al., 2007).

Discipline. The term discipline was used as an independent variable to identify the number of occurrences of behavior for which a disciplinary or behavioral log entry was recorded for a student, regardless of severity (Balfanz, Bridgeland, Moore, & Fox, 2010).

Dropout. The term dropout was the central term of this entire study, and throughout other studies was given the most varied definition of any of the key terms. The definition of dropout for this study most closely resembled the definition used by

Alexander et al. (1997), with one modification because the state in which the district in this study was located recognized a student receiving a General Equivalency Diploma (GED) as a dropout. The term dropout was defined in this study as a student who was not enrolled and had not graduated from high school with a diploma at the time the data were collected.

Family variables. The term family variables was used in this study to refer to student risk factors for which the school had little or no involvement. While previous research had provided numerous examples of variables of this type, this study only used socioeconomic status as a family variable (Gleason & Dynarski, 2002).

Free or reduced lunch status. Free or reduced lunch is a federally defined status that is often used to determine a student's socioeconomic status. It is not affected by local district decisions. Maximum household income levels are set each year for students to qualify for free or reduced lunch prices at school (Missouri Department of Elementary and Secondary Education, 2012).

Grade retention. Grade retention was used to refer to a student who had to repeat and complete an entire grade in school (Kennelly & Monrad, 2007).

Grades. Grades were measured as a final grade point average in core classes. Core classes included mathematics, science, social studies, or English classes. Grade point average was calculated on a four point scale, with an A earning four points, a B earning three points, a C earning two points, a D earning one point, and an F earning zero points. All core grades were then averaged to develop a value between 0.0 and 4.0 for each student (Balfanz et al., 2007).

Intervention. Interventions were defined as any efforts schools made to change the course on which a student was headed. For this study, interventions were typically efforts to keep students from dropping out of school.

Mobility rate. Mobility rate was not used as a variable in this study due to the limitations of the data collection process and following students who change schools. The term mobility rate was used, however, to discuss a risk factor of dropping out signaled by frequent changing of schools (Rumberger & Larson, 1998).

Pull-out factors. Risk factors that lead to higher incidence of dropping out of school generally fall into two categories: pull-out factors and push-out factors. Pull-out factors refer to factors outside of the school that may entice a student to leave school (Stearns & Glennie, 2006). Examples of pull-out factors include the job market, family responsibilities, and peer relationships. Pull-out factors were discussed in this study but were not used as predictor variables due to the nature of the model being developed.

Push-out factors. Push-out factors differ from pull-out factors in that push-out factors refer to factors controlled by the school. These factors may lead students to leave school to remove themselves from the effects of the factors (Stearns & Glennie, 2006). Examples of push-out factors include disciplinary consequences, grading policies, teacher/student interactions, and building climate. The push-out factors being considered as variables in this study were: attendance, grades, standardized test scores, discipline, reading level, and grade retention.

School variables. For this study, school variables referred to risk factors that generally occurred in or were affected by the school. Examples of school variables used in this study included grades, attendance, and discipline (Balfanz et al., 2007).

Socioeconomic status. Socioeconomic status is an independent variable that was measured by whether or not a student qualified for free or reduced lunch during a given school year (Montes & Lehmann, 2004). Free or reduced lunch is a federally defined status and is not affected by local districts. Free lunch and reduced lunch were combined for reporting purposes, so students were reported as either full price lunch or free/reduced lunch.

Significance of the Study

The significance of this study was based on three types of contributions: contributions to the literature, contributions to practice, and contributions to the author's institution. As it pertains to the literature, the study was placed in a lightly-researched area. The literature is rich with dropout prevention and identification strategies at the high school level (Heppen & Therriault, 2008; Jerald, 2006; McKee, Melvin, Ditoro, & McKee, 1998) and to a lesser extent at the middle school level (Balfanz et al., 2007; Bowers, 2010; Rumberger, 2007). There is little research centered on elementary-age predictors (Alexander et al., 1997; Annie E. Casey Foundation, 2010; Jimerson et al., 2001; Montes & Lehmann, 2004). If this research study successfully identifies a prediction model, it will enhance a minimally-researched field. If it does not successfully build a model, it will add more complete results to the little existing elementary research.

As it pertains to practice, this study should improve effectiveness and efficiency. The goal was to build a model that was very practical and based on school-generated data. Other prediction models rely at least partially on surveys completed by parents and school staff. These are time consuming and impractical to complete on a large scale, so many potential dropouts may be missed (Balfanz et al., 2007). If an easier-to-use model could be developed, it would help schools routinely look at all students to identify potential dropouts. This would improve the use of interventions beginning at a much younger age.

As it pertains to the author’s institution, the significance of the study is the same as it is for other institutions. This research will make very practical contributions to the author’s institution, much like those outlined above. It is hoped that a very practical approach springs from this research, enabling the author’s institution and many others to utilize this work to take dropout prevention strategies to a more successful level.

Summary

Dropping out of school has been shown to have many potential consequences for the students involved. As a result, school officials have put much effort into dropout prevention. Because it is often too late to prevent dropouts once students reach high school, attention has turned toward earlier identification of students at risk of dropping out of school. In this way educators hope to design interventions to help the students before they ever reach the levels of wanting to drop out of school.

In order to successfully predict students at risk of dropping out, many models have been developed to use as identification tools for schools. Unfortunately, most of the existing tools are lacking in some fashion. This study intended to fill that gap by developing a model that was early, accurate, and practical enough for schools to use.

Students tend to develop feelings, positive or negatively, about school in the elementary years that stay consistent throughout the rest of their schooling. Combine this with the fact that many different factors can lead to the decision to drop out for different students, and the challenge becomes even greater. If an early, accurate, and practical model could be developed for predicting which students are at risk of dropping out, schools will have a greatly increased chance at successfully intervening for those students.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

The first chapter presented a brief background of the dropout problem, including a discussion of the magnitude of the problem and what school leaders have tried to do to help. Various authors offered risks that dropouts face later in life, providing motivation to reduce the number of students who drop out of school. This led to the purpose of the current study, which was to identify potential dropouts at an early age so interventions could have a greater chance of being successful at keeping students in school through graduation. This chapter will synthesize literature on the topic.

A review of the relevant literature revealed several main topics that contributed to the development of this study. These topics will be presented in the sections that follow, beginning with a general background and description of the problem at the center of the study. This includes background on how pervasive the dropout problem is in America as well as the lifelong risks faced by students who choose to drop out of school prior to graduation.

The second section will examine what schools have typically done to intervene for students deemed at risk of dropping out of school. This will include a description of intervention programs and the kinds of characteristics they have attempted to address to prevent dropouts. The third section of the literature review will look closely at why interventions have typically not been successful at reducing the national dropout rate. Many different factors can contribute to the success or failure of these efforts, and these factors will be discussed.

The fourth section of the review will discuss why earlier intervention efforts are more successful than later efforts. This section will also cover the loss in accuracy that typically accompanies earlier identification of students in need of intervention. The fifth section in this chapter will detail current and former attempts at early identification of students at risk of dropping out of school. These attempts have all had shortcomings of one kind or another, and these limitations will be discussed.

The final section of the literature review will cover previous attempts at developing prediction models specifically using regression techniques, which is what the current study intended to do. The techniques of several studies will be discussed, including their findings and how those studies influenced the current study. All of these sections will lead to the motivation for the current study, which is to accurately identify potential dropouts as early as fourth grade in order to give interventions a better chance of being successful at keeping students in school through high school graduation.

Background of the Dropout Problem

This section will describe the background of the dropout problem. It will include estimates of the extent of the dropout problem. Following that, this section will discuss the kinds of negative characteristics that surface later in life at much higher rates for dropouts than for high school graduates. This discussion will include personal, societal, and other characteristics and will reveal why school leaders feel a sense of urgency to address the dropout problem.

Estimates of the National Dropout Rate

Due to different definitions of dropouts and different reporting methods, it is difficult to identify exact rates of students dropping out of school prior to graduation.

Recent research does not even agree on whether the dropout rate has improved or gotten worse than in past years. Neild, Balfanz, and Herzog (2007) stated dropout rates are generally as good as they have ever been, yet Rumberger (2011) noted dropout rates appear to have gotten worse, leaving them higher than they were 40 years ago.

Despite the difficulties in pinning down exact numbers, many researchers pointed to dropout rate estimates as motivations for their studies. The most optimistic estimates of the dropout rate tended to hover around 25% (Bowers, 2010; John W. Gardner Center, 2011; Rumberger, 2007). Others placed the estimate in the vicinity of 30% (Kennelly & Monrad, 2007; Rumberger, 2011). Some estimates shot even higher. Bowers (2010) pointed to some estimates over 30%, while Heppen and Therriault (2008) stated almost one-third of students drop out prior to high school graduation.

While these estimates of the dropout rate do not differ by drastic margins, two main points are worth noting. First, since there is not a definitive dropout rate agreed upon by all, it is evident data reporting in this area is not accurate or consistent enough to settle on universal national measurements. Second, regardless of which estimate is used, the dropout rate for American students is alarmingly high. To consider that at least one out of every four students fails to graduate from high school is cause for alarm.

Personal Concerns That Tend to Surface for Dropouts

Moving beyond estimates of the extent of the problem, most rationales for studying the dropout problem pointed to the increased risks of concerns faced later in life by people who chose to drop out of school before graduating. One of the most common concerns listed was the higher likelihood of reduced income and lower lifetime earnings (Alexander et al., 1997; Annie E. Casey Foundation, 2010; Kennelly & Monrad, 2007;

Rumberger, 2011; Zvoch, 2006). A couple of researchers stated this issue more elaborately. Christenson and Thurlow (2004) stated, “Jobs that pay living wages and benefits have virtually disappeared for youth without a high school diploma” (p. 36). Neild, Balfanz, and Herzog (2007) similarly stated, “It is practically impossible for individuals lacking a high school diploma to earn a living or participate meaningfully in civic life” (p. 28).

Societal Concerns That Tend to Surface for Dropouts

A related concern frequently referred to by researchers was the increased financial burden placed on society by dropouts who fail to make the wages they might have otherwise earned with a high school diploma. This included lost revenue from taxes (Annie E. Casey Foundation, 2010; Kennelly & Monrad, 2007), but more significantly, it meant higher costs paid out through welfare programs for some of those individuals (Alexander et al., 1997; Christenson & Thurlow, 2004; Rumberger, 2011; Zvoch, 2006).

Another concern of dropping out of school prior to graduation was an increase in the rates of criminal activity and eventual imprisonment over the same rates for high school graduates (Alexander et al., 1997; Annie E. Casey Foundation, 2010; Christenson & Thurlow, 2004; Rumberger, 2011). This was significant not only because of the financial costs to society but also because of the stresses added to families and individuals.

Additional Concerns That Tend to Surface for Dropouts

In addition to the problems mentioned frequently by numerous researchers, there were also several other lifetime issues noted by some researchers that appear more frequently for dropouts than for high school graduates. These additional concerns were not routinely listed as major issues by authors, yet they were mentioned in some research. These other concerns included increased risks of teenage childbirth (Annie E. Casey Foundation, 2010), general lack of productivity to society (Annie E. Casey Foundation, 2010; Kennelly & Monrad, 2007), higher unemployment rates (Christenson & Thurlow, 2004; Zvoch, 2006), and poorer health and increased mortality (Rumberger, 2011).

Because of the increased risk of serious concerns faced later in life by people who drop out of school before graduating, educators have made concerted efforts to identify and intervene for students deemed at-risk of dropping out of school. These efforts have taken a variety of forms, but the main focus has been to try to reduce the dropout rate so more students are not subjected to the increased risks of negative concerns associated with dropping out of school.

Intervention Efforts

This section will begin by discussing how schools have historically viewed the dropout problem. It will then move into a summary of intervention programs that have been used in an attempt to reduce dropout rates. This will rely heavily on a compilation of intervention programs provided by Kennelly and Monrad (2007). This section will also discuss how some different programs focused their efforts, which will lay the foundation for the next section about why many of these programs have not worked.

Historical Viewpoint of Dropout Problem

Students dropping out of school prior to graduation is not a new phenomenon. According to Schargel and Smink (2001), studies have indicated a dropout rate as high as 90% in 1900. The same authors also noted that in 1945, the economy could absorb dropouts, meaning there were available jobs at living wages for them. As society changed, that became untrue, and schools began trying to intervene to keep all students in school through graduation.

Dynarski and Gleason (1998) discussed the fact that school districts have been operating dropout prevention programs for many years, but it was not common for districts to evaluate the effectiveness of such programs. Schools often relied on observational data and anecdotal information to identify potential dropouts, and programs to intervene were not research based. Efforts were not targeted to specific factors nor were they adjusted over time based on evidence of effectiveness or the lack thereof.

Intervention Programs

In more recent years, schools, districts, states, and the nation have begun implementing intervention programs that are more targeted to specific deficiencies of students. Kennelly and Monrad (2007) compiled a list of 26 intervention programs adopted by schools to help address dropout rates, including Achievement for Latinos Through Academic Success, Career Academies, Check and Connect, Positive Behavioral Interventions and Supports, and RTI. For each of these programs, they looked at 15 characteristics to determine which areas each program addressed.

Kennelly and Monrad (2007) did not categorize their characteristics, but in order to describe them more easily, they can naturally be grouped into four areas: academic help; general student support and relationship building; general support for students in and outside of school; and academic and behavioral support. Some of the topics frequently addressed by intervention programs were in the area of academic help. In this area, Kennelly and Monrad looked at the following characteristics: focus on achievement in core courses, tutoring as an academic support, and catch-up courses. These were all characteristics of programs aimed at improving course grades or other course-specific outcomes.

Another area looked at by Kennelly and Monrad (2007) was general student support and relationship building. In this area they listed the following characteristics: Counseling/Mentoring; Small learning communities for greater personalization/School within a school; Homeroom, teams or looping; and Ninth Grade Academies or transition programs. These characteristics were all focused on helping students adjust to school and to be comfortable at school.

A third category of characteristics of intervention programs could be described as general overall support for students, both in and out of school. This area included the following characteristics: focus on positive effects for diverse students; focus on positive effects for students with disabilities; career/college awareness; family engagement; community engagement; and partnerships between high schools and feeder middle schools (Kennelly & Monrad, 2007). These characteristics highlight the notion that many intervention programs for students did not just look simply at grades or other singular factors but strove to offer students a whole system of support.

The last two characteristics did not seem to fit in the other categories but defined an academic and behavioral support category. In this area, Kennelly and Monrad (2007) listed the following characteristics: attendance and behavior monitors and a tiered approach to providing behavioral and/or academic support from universal to most intensive. These two characteristics drew attention to two of the factors most often considered in school settings – grades and discipline. As is shown elsewhere in this study, these are two of the most commonly listed predictors of dropouts.

A close look at the 26 intervention programs studied by Kennelly and Monrad (2007) revealed the programs ranged from specifically targeting just one characteristic to encompassing 6 of the 15, meaning none of the programs attempted to address even half of the characteristics noted. Six of the programs targeted just one of the characteristics listed, two of the programs encompassed 5 of the 15 characteristics, and one program addressed 6.

Simply attempting to address more areas did not necessarily make programs more successful. In fact, Kennelly and Monrad (2007) noted, “Few programs have demonstrated positive (or potentially positive) effects” (p. 12). They discussed that only 3 of the 26 programs considered did show positive or potentially positive effects: Achievement for Latinos Through Academic Success, Career Academies, and Check and Connect. Those three programs addressed 4, 4, and 5 of the 15 characteristics, but had only one characteristic in common: community engagement. This highlighted the idea that preventing dropouts can be a larger scale problem than simply addressing issues inside the school building.

The next section will take a closer look at why most of these intervention programs have not been successful. School officials have started trying to implement packaged programs to tailor to their needs, but they have not necessarily experienced much success. This can occur for a variety of reasons, and those will be discussed in more detail in the following section.

Why Interventions Have Not Worked

As discussed in the previous section, Kennelly and Monrad (2007) looked closely at a number of intervention programs and broke them down into the different characteristics addressed with students when trying to prevent them from dropping out of school. Christenson and Thurlow (2004) supported this notion. They suggested interventions should be targeted to the specific deficiencies students experience, and these kinds of specified interventions can show some success at preventing dropouts. It should be noted that they referred to experiencing some success, but not complete success. In reality, most intervention efforts have not realized the success rates those implementing them envisioned. Many different reasons have contributed to that shortcoming for different programs.

Practicality of Identification Methods

One key factor that can affect the success of intervention efforts is whether or not the process used to identify students is practical enough to be used broadly, thereby enabling a school to consider all students when attempting to identify those in need of help. Alexander et al. (1997) conducted a very thorough study using many variables to try to identify students at risk of dropping out of school. Some of the variables in their study were grouped into a category they called “Family Context” (p. 88). This category included variables such as family stressors, parents’ attitudes and values, and parents’ socialization practices. While these may have been very helpful in identifying potential dropouts, these kinds of variables were impossible to collect from school data. The only way to get this information was to conduct a survey of the family. It would not be feasible for a school or district to conduct this kind of survey with all families in the school or district, meaning this data would not even be collected unless there was first some indicator that led to this family in the first place.

Caraway et al. (2003) had a similar situation in their research. They included in their prediction model a category of self variables. This included variables such as self-efficacy, goal orientation, test anxiety, and general fear of failure. They collected this data by administering separate self-report questionnaires for each variable, ranging from 4 items to 37 items, for a total of 79 items. Additionally, as with the Alexander et al. (1997) study, these variables were only one portion of the overall prediction tool. The same problem would surface for Caraway et al. (2003). Collecting this data on all students would be impractical, meaning other measures would need to be used to identify potential students in the first place. Often these other measures are observational and subjective, meaning there would be a high likelihood that students at risk of dropping out would never even be identified to get the questionnaires, much less the intended interventions.

Another example of the difficulties faced by many studies of dropout predictors was presented by Gleason and Dynarski (2002). They referred to commonly stated categories of variables shown through research to be associated with dropping out of school. These included personal/psychological characteristics, adult responsibilities, and school or neighborhood characteristics. In addition to the need to collect survey information from students and parents, this would also require schools or districts to collect specific information about the local neighborhood for each identified student. This again would cause difficulties for schools in finding all potential dropouts.

This practicality issue was significantly addressed by Balfanz et al. (2007) when they stated:

> We are skeptical, however, that such an approach will ever be common in district-based dropout prevention programs. It is uncommon for districts to routinely

> administer extensive surveys to all their students or to have the expertise or leisure to create valid, reliable, and highly predictive scales and then to use these scales in sophisticated cluster analyses to classify all their students into various categories of risk. (p. 225)

They felt that for identification techniques to be truly useable, they must be practical enough to be used on a large scale, meaning they could easily be applied to an entire school or grade level.

Accuracy of Identification

A second key factor that has affected the success of intervention programs is the accuracy of the identification tools used. Certainly, dropout intervention efforts cannot be successful if they are not implemented for the right students. Failing to identify students in need of intensive support can quickly contribute to a reduced impact of any intervention program. Falsely identifying students as at-risk who would graduate with no intervention is also problematic. This causes schools and districts to misdirect precious intervention resources toward students who do not need them, leaving reduced resources to help students who truly need support.

Gleason and Dynarski (2002) claimed misidentification of students in need of intervention was actually a quite common problem. They stated that many intervention efforts are aimed at the wrong students. In order for identification efforts to be truly accurate, they must not fail to name students actually at-risk nor can they falsely label too many students as at-risk who are not. It is extremely difficult if not impossible to identify future dropouts perfectly. Every research study reviewed for this paper discussed problems with identification in both directions – missing some students actually at-risk of dropping out and falsely labeling others who were not.

Timing of Interventions

A third key factor that has affected the success of intervention programs is how far along the targeted students are in their education. Because high school personnel often feel the most direct pressure to reduce dropout rates, many intervention programs have been implemented at the high school level. A positive finding that came out of several research studies was that, when utilized at the high school level, prediction instruments can identify students at risk of dropping out quite well (Battin-Pearson et al., 2000; Caraway et al., 2003; Heppen & Therriault, 2008).

Unfortunately, waiting until high school to identify students at-risk of dropping out has been shown to be problematic for another reason. Bowers (2010) argued that high school is much too late to begin trying to intervene for potential dropouts. He discussed that the negative effects these students experience accumulate over time, rendering later intervention efforts less effective. Suh et al. (2007) similarly stated that if interventions are not implemented until problems have been evident for a time, more intensive efforts are needed to intervene, and even then the impact of intervention programs will be reduced.

In order to address this problem with interventions starting too late, researchers have tried to develop identification tools that can draw attention to students at-risk of dropping out at younger ages (Alexander et al., 1997; Battin-Pearson et al., 2000; Bowers, 2010; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). If earlier identification were possible, interventions could start earlier. The next section will address some of the factors involved in earlier identification.

Factors Involved in Earlier Identification

As stated in previous sections, earlier identification of potential dropouts can be desirable in order to give interventions an increased chance of being effective. This section will discuss some of the research that supports the need for earlier identification. Following that, this section will also discuss the decrease in accuracy of early identification methods as compared to methods used when students are older.

Need for Earlier Identification

As discussed in the previous section, one of the major barriers to effectively intervening to prevent students from dropping out is the timing of the intervention. Suh et al. (2007) discussed the importance of earlier identification. They noted that often the effectiveness of interventions is weakened once multiple risk factors have appeared. To combat this problem, they suggested schools look for identification models that allow them to begin interventions earlier in the educational process.

Christenson and Thurlow (2004) considered how students get to the point of dropping out over time, stating, “Early and sustained intervention is integral to the success of students because the decision to leave school without graduating is not an instantaneous one, but rather a process that occurs over many years” (p. 37). This idea became central to the argument of Alexander et al. (1997) when they noted children typically begin school with a positive attitude and an excitement for learning. Waiting to provide interventions for students until their attitudes toward school begin to “spiral downward” (p. 87) creates difficulty in ultimately keeping them in school through graduation.

Bowers (2010) similarly pointed to the long-term nature of the decision to drop out of school. He noted the decision to drop out is not generally based on a single factor, but it more often stems from long-term effects of multiple factors. He argued that school leaders should start identifying at-risk students earlier, stating, “For many districts nationwide, early student dropout identification is critically important so that the district can potentially intervene early in a student’s schooling career to help delay or prevent dropout” (p.192). Bowers continued by discussing that early preventative services would be more successful than waiting until much later to try to motivate a potential dropout to change his or her mind.

Accuracy of Earlier Identification

While Bowers (2010) supported the need for earlier identification of potential dropouts, he also presented a problematic issue. He noted that middle school identification was less accurate than high school identification of future dropouts. Gleason and Dynarski (2002) went further with this, looking at over 20 potential risk factors at the middle school and high school levels. Among others, these factors included: high absenteeism, over-age for grade level, single parent homes, failure to do homework, and low self-esteem. They also considered multiple risk factors present for the same student. For each variable they considered what percentage of students with that characteristic actually dropped out of school.

Upon examining their data, Gleason and Dynarski (2002) reported that, without fail, each factor presented a higher dropout rate when present in high school students than in middle school students. Even more telling is the specific rates they shared for their regression model built from all factors. Gleason and Dynarski (2002) reported a dropout rate of 42% when identified by their regression model for high school students, but only 23% of identified middle school students eventually dropped out of school. This showed a significant drop in accuracy of their identification model when moving earlier from high school identification to middle school identification.

This led to the problem central to the current study. Authors including Suh et al. (2007), Christenson and Thurlow (2004), and Alexander et al. (1997) discussed the need for earlier interventions, which necessitated the need for earlier identification. Other authors, including Bowers (2010) and Gleason and Dynarski (2002), noted the difficulty in accurately identifying these students earlier. Many authors have attempted to solve that problem by finding accurate ways to identify potential dropouts earlier. These authors will be discussed in the next section detailing previous attempts at early identification of future dropouts.

Early Dropout Identification Attempts

Previous sections have provided rationale for why potential dropout identification efforts have continued to focus on earlier and earlier ages. This section will highlight some of the attempts at identifying dropouts at various stages in their educational careers. The discussion will begin with efforts to identify dropouts during high school years, which have tended to be relatively accurate. The section will continue by addressing efforts focused on middle school-aged students, which have typically been somewhat less accurate than high school identification. Finally, this section will cover efforts to identify students as early as elementary school, which have generally been even less effective.

Identifying Potential Dropouts During High School

Most students who drop out of school do so during high school (Jerald, 2006), so some researchers have focused identification efforts on the high school years. For interventions to be applied, the appropriate students must first be identified. McKee et al. (1998) addressed the problem by attempting to develop a scale that could be applied to a range of students in order to identify those at-risk of dropping out and to determine which specific areas caused the most significant concern for each student.

The first scale developed by McKee et al. (1998) was the SARIS-AQ, an administrator questionnaire containing 13 different measurements generally available in school records, including attendance, discipline, and grades. Ideally, the scale could be used on all students in a particular grade or school to identify those most likely to eventually drop out of school. In testing their scale, McKee et al. (1998) found that the scale correctly predicted the eventual dropout or graduation status correctly for 84% of the students considered. While this is an extremely high percentage to correctly predict, their work presented several limitations. First and foremost, as discussed earlier in this study, the identification did not occur until high school, rendering intervention efforts less effective than earlier efforts potentially could provide (Alexander et al., 1997; Bowers, 2010; Christenson & Thurlow, 2004). Second, their scale was tested on a sample of only 49 students, which is too small a sample to develop definitive conclusions (Field, 2009). Third, the authors concluded that many schools would not have the resources to apply the scale to entire populations, meaning some students would not be identified as needing interventions by the scale.

McKee et al. (1998) addressed the last limitation by developing a different scale, the SARIS-SQ, which was a questionnaire completed by students. This made it easier to apply to large populations to ensure all students would be considered. Their work with this scale covered 15 questions and showed a 94% rate of correctly predicting dropout or graduation status. While the SARIS-SQ was used on a wider scale than the SARIS-AQ and tested on a much larger sample of 423 students, it still had limitations. First, the identification did not occur until high school. Perhaps the greater limitation, however, is the difficulty in generalizing the scale to other settings. McKee et al. (1998) developed a cut score for the scale, meaning that students who scored at or higher than that cut score would be considered at-risk of dropping out of school. The problem was that the cut score was developed in their study with the knowledge of who did and did not drop out of school. The study suggested schools should set their own cut scores to determine who needs interventions, and they discussed that the cut score could vary. It would be difficult for schools to rely on such a measurement, since an accurate cut score could only be determined by already knowing who graduated and who dropped out of school. A positive characteristic of the scale is that, regardless of cut score, it is designed so higher scores generally indicated higher risk of dropping out of school. With this in mind, schools could at least use the scale to give some idea of which students need interventions more than others, even if they cannot necessarily predict which students may drop out without interventions.

Another example of a study focusing on high school students was reported by Heppen and Therriault (2008). While they acknowledged the range of predictor variables reported by other researchers, they chose to focus their work on just two main variables:

attendance and grades. They intended to build a system of identification that would be feasible and easy to apply to all students in a setting, and they recommended ninth grade as an important milestone year to track.

Heppen and Therriault (2008) determined from other research that missing more than 10% of school days should be considered a cause for concern. They also determined that falling off-track for graduation, meaning a student was not on pace to graduate after ninth grade, was a major indicator of a student being at-risk of dropping out of school. The authors suggested schools could easily track these two categories of data for all students and provide interventions for those who seem to need them.

A positive aspect of this study was that Heppen and Therriault (2008) worked to keep their process practical for schools to apply to all students. In doing this, however, there were still two significant limitations. First, as with other studies focusing on high school students, any interventions offered for identified students will likely be less effective than if they were offered earlier (Alexander et al., 1997; Bowers, 2010; Christenson & Thurlow, 2004). Second, in their attempt to keep the process simple and practical, Heppen and Therriault (2008) did not utilize several variables repeatedly found by other researchers to be indicators of potential dropouts, including discipline, test scores, and socioeconomic status (Battin-Pearson et al., 2000; Gleason & Dynarski, 2002; Rumberger, 2007).

As indicated for these studies, researchers choosing to focus identification efforts on high school students are all plagued by the same limitation – that waiting until high school to identify and intervene for potential dropouts is too late to be fully effective (Alexander et al., 1997; Bowers, 2010; Christenson & Thurlow, 2004). In an effort to improve identification and intervention efforts, some researchers began to focus on middle school students. Some of those efforts will be discussed in the next section.

Identifying Potential Dropouts During Middle School

Rumberger (2007) attempted to develop a method for earlier identification of potential dropouts, focusing his attention on middle school students. He looked at two main areas of risk factors: demographic factors and student performance measures. For demographic factors, Rumberger (2007) focused on eighth graders and considered the following characteristics: single-parent households; parents who did not graduate from high school; older siblings dropped out of school; spending three or more hours at home alone each day; limited English proficiency; and low socioeconomic status. Using these characteristics, Rumberger (2007) found that students possessing three or more of the risk factors only graduated about 50% of the time. While this can still be a good tool for schools to use, one limitation is that this prediction percentage falls below many of those reported for high school-age identification strategies, as expected from the work of Bowers (2010).

For student performance measures, Rumberger (2007) focused on sixth graders and considered the following risk factors: failed English; failed math; unsatisfactory behavior; and attendance rate of 80% or less. He found that, for his sample, 71% of students with at least one risk factor dropped out before graduating. He went on, however, to list some of his own limitations. First, he noted 41% of dropouts did not possess any of the risk factors, meaning they would not have been identified as needing interventions by his model. He also discussed that some of the students who did possess risk factors graduated without any interventions, meaning potentially using resources to intervene for them would have meant using resources unnecessarily.

Balfanz et al. (2007) attempted to address the limitations found in these other studies. They focused their research on sixth grade students in an attempt to identify potential dropouts earlier than some other studies so that interventions would have a better chance of being successful (Alexander et al., 1997; Christenson & Thurlow, 2004; Suh et al., 2007). Balfanz et al. (2007) also limited their risk factors to those commonly available in school reports. This enabled them to keep their model practical enough to be applied to students on a large scale, as opposed to those models that required personal surveys, interviews, or questionnaires with students, teachers, or family members.

Through their analysis, Balfanz et al. (2007) identified five variables that showed predictive power for eventual dropouts: attendance rate of 80% or less; failing math in sixth grade; failing English in sixth grade; receiving an out-of school suspension in sixth grade; and getting an unsatisfactory final behavior mark in any subject for sixth grade. Students with these risk factors present in sixth grade failed to graduate at very high rates. Students with unsatisfactory behavior marks dropped out at a rate of 71%, and all four of the other factors yielded dropout rates of 80% or higher. As with other studies, even though these rates are very high, there were still some students who dropped out without showing any of these factors in sixth grade. This led to an overall prediction rate of 60% for the model used by Balfanz et al. (2007).

Having responded to the need for earlier identification, practicality of the model, and a relatively strong ability to predict dropouts, the study by Balfanz et al. (2007) presented a very solid model for identifying potential dropouts earlier than high school so interventions could be applied at younger ages. While they addressed many of the limitations of other studies, there was still one limitation for this study: some researchers have determined that students who drop out of school often began moving down the path toward dropping out during elementary school (Alexander et al., 1997; Bowers, 2010; Christenson & Thurlow, 2004).

While efforts to identify potential dropouts in middle school have attempted to improve on models for which identification did not happen until high school, there is still a need to try for even earlier identification. For this reason, some researchers pursued the development of identification models focused on elementary students. The next section will discuss some of these efforts.

Identifying Potential Dropouts During Elementary School

Much of the research has focused on high school and middle school predictors of dropouts, and elementary predictor research is not as abundant. For those researchers who have considered elementary predictors, they have generally not developed prediction models focused solely on those ages, but they have included elementary predictors in a general discussion along with middle and high school factors (Alexander et al., 1997; Jimerson et al., 2001; Montes & Lehmann, 2004).

In their discussion, Montes and Lehmann (2004) developed a list of predictors of dropouts at various ages based on their review of literature. They started before elementary school, listing quality of care giving as a predictor. They continued with first grade and later elementary predictors, including problem behaviors, school performance, grade retention, parent involvement, gender, socioeconomic status, stressful life events, and mobility. Montes and Lehmann (2004) listed all the same predictors for middle and high school, but added absenteeism, disciplinary problems, and a self-report on graduation likelihood.

Montes and Lehmann (2004) expanded their discussion on some of these factors to include how well they predicted dropouts. They reported that being retained in first grade led to a 300% increase in likelihood of dropping out later. They also found that first graders with multiple risk factors dropped out 80% of the time, and students retained in elementary and middle school dropped out 94% of the time. Montes and Lehmann (2004) did not develop a tool for identifying potential dropouts, but they listed key factors that schools could monitor and use to help themselves identify students in need of intervention at all grade levels.

Another report looked at one specific elementary predictor of dropping out of school. The Annie E. Casey Foundation (2010) reported that reading proficiency by the end of third grade was a strong indicator of future dropouts. Specifically, the authors noted, "millions of American children get to fourth grade without learning to read proficiently. And that puts them on the dropout track" (p. 7) and "A person who is not at least a modestly skilled reader by that time is unlikely to graduate from high school" (p. 9). Their reasoning behind these statements was that through third grade students were generally learning the mechanics of how to read, but after that they were using reading to learn other topics. If a student did not master the mechanics of reading by that time, often they were left behind after the classroom focus on learning how to read had shifted to other topics with the assumption that students had grasped basic reading skills.

While the report from the Annie E. Casey Foundation (2010) discussed reading proficiency in elementary school in great detail, the purpose of the report was not to discuss other predictors of future dropouts nor was it to specifically use reading proficiency to predict future dropouts. The idea of focusing on one specific factor, though, served to bring attention to one topic that schools could attempt to affect for students who struggle.

Both the Annie E. Casey Foundation (2010) and Montes and Lehmann (2004) discussed elementary predictors of dropouts. As identification tools, however, they were both lacking. Neither collected data from a sample of students to use data analysis to support their claims or to offer more detailed information. Both reports did attempt to provide motivation and encouragement to consider elementary factors as dropout predictors, responding to the limitations of other studies where the identification of at-risk students came later. The lack of an identification tool led to the need for the current study, which attempted to develop a quantitative identification tool based on school-based data.

Previous Identification Models Developed Using Regression

The current study intended to use multiple regression to develop an identification model to predict future high school dropouts during elementary school. Among the studies previously discussed, four of the authors used regression in their studies to develop their prediction models (Balfanz et al., 2007; Bowers, 2010; Suh et al., 2007; Uekawa, Merola, Fernandez, & Porowski, 2010). Their studies differed from each other, and as a result their regression attempts produced different findings. The current study drew from these studies to help develop the methodology for data analysis. The findings of those authors will be discussed in this section.

Suh et al. (2007) used logistic regression to analyze their data. They began by considering 20 different potential dropout predictors. Of these 20 predictors, 8 of them were available through school-generated reports, while the remaining 12 had to be collected through surveys. To begin to understand the effect of the different predictor variables on eventual dropout rates, Suh et al. (2007) used Pearson correlations to draw initial conclusions about the effects of each variable.

The data used by Suh et al. (2007) were from the National Longitudinal Survey of Youth 1997 but, while the study collected data from a range of ages, they chose to focus mostly on the presence of predictors at the eighth grade level. The Pearson correlation revealed eight variables that showed stronger correlations with dropping out than with the other predictor variables. These eight characteristics were: low grade point average, suspensions, living with both biological parents, low socioeconomic status, number of schools attended, percentage of peers going to college, first sexual experience prior to age 15, and optimism about the future.

Further analysis of the data was then done by Suh et al. (2007) using logistic regression to control for the relationships between variables. Through this analysis, they determined that 14 of the 20 variables they studied were statistically significant predictors of dropouts. Suh et al. (2007) noted three of the strongest predictors of dropouts in their study were commonly mentioned in other research: low GPA, suspensions, and low socioeconomic status. In addition, they also found two other variables to be strong predictors of dropouts: first sexual experience prior to age 15 and highest educational attainment of the mother being high school or less.

The conclusion drawn by Suh et al. (2007) was that eventual dropout or graduation status was correctly predicted by their model for almost 82% of the adolescents in the study. This was a high rate of accuracy, but as pointed out earlier in this study, the identification of potential dropouts based on eighth grade data is often too late. The study conducted by Suh et al. (2007) influenced the current study by looking deeper at the effects each variable had on dropping out, given that many other studies just calculated simple percentages of students possessing each variable who dropped out of school. One aspect of the study by Suh et al. (2007) that was not replicated by the current study was the inclusion of variables that had to be collected through surveys. The current study intended instead to use only variables widely available through school-generated reports.

Another group of authors who used logistic regression to develop a dropout prediction model was Uekawa et al. (2010). Similar to other studies, Uekawa et al. started with a list of predictor variables commonly mentioned in other studies. They initially considered attendance, behavior, final course grades for English and math, charter school versus regular school, race, socioeconomic status, English language learner status, special education status, gender, and grade retention.

Uekawa et al. (2010) discussed the importance of developing a model using as few variables as possible. Given that dropping out of school is a relatively rare occurrence statistically, they concluded that a model using a long list of predictors could not be statistically supported. Thus, while many of their predictor variables showed independent associations with dropping out of school, they refined their model to determine the strongest indicators when controlling for other indicators.

Through their regression analysis, Uekawa et al. (2010) determined that three variables showed high levels of consistency as predictors of dropout status. These variables were attendance, grade retention, and final course grades for English and math. They also specifically discussed their findings relative to the behavior variable. While they found it to be true that dropouts had higher overall incidence of behavior problems, none of their behavioral measures showed statistically significant predictive powers relative to other variables.

In assessing the accuracy of their prediction model, Uekawa et al. (2010) noted that the model successfully identified 58% of eventual dropouts as at risk of dropping out of school. On the other hand, their model labeled a large number of students as at risk of dropping out, and only about 12% of those identified actually dropped out of school. Because their model incorrectly identified so many students and still correctly identified just over half of all dropouts, Uekawa et al. (2010) described their results as a “large prediction failure rate” (p. 11). They went on, however, to discuss that such a result was inevitable due to the relatively low occurrence of dropouts.

The study conducted by Uekawa et al. (2010) was useful to the current study because it included many of the same variables as the current study considered. Their explanation of findings helped inform the current study by reiterating the need to find the strongest predictor variables even though all considered variables may show some relationship to dropout status. Additionally, the lack of accuracy in their findings was a motivating factor for the current study to analyze the variables by a slightly different method.

Bowers (2010) employed a discrete-time hazard analysis using logistic regression to develop his dropout prediction model. By using this method, he was able to account for the fact that once a student drops out, no matter how early, that student's generation of additional data ends. He wanted to account for all years of each student's education without affecting his data by including characteristics that were no longer part of the cohort's makeup.

Through his review of literature, Bowers (2010) chose to consider seven variables as potential predictors of dropouts: time in school (measured in years), gender, ethnicity, district attended, total occurrence of D or F letter grades, grade retention (whether a student was ever retained during his or her educational career), and GPA. Bowers used three of these variables as dichotomous variables: gender, district, and ethnicity. District was dichotomous because Bowers collected data from students in two different districts. Ethnicity was treated as a dichotomous variable because Bowers differentiated between European Americans and all other ethnicities. These three variables were not time-variant.

Since Bowers (2010) was considering effects over time, he included three variables that were time-variant: grade retention, total occurrence of D or F letter grades, and GPA. He felt it was important to note that each of these variables had the potential to produce different measurements for each student based on which year in school was considered. Bowers used the discrete-time hazard model with logistical regression to analyze his data because he was interested in both when students were predicted to drop out as well as which variables seemed to best indicate that risk in the first place.

Through his regression analysis, Bowers (2010) was able to explain over 50% of the variance in the probability of a student dropping out of school. He felt this was an improvement over previous attempts at building prediction models, but he noticed that none of time-invariant variables were included in the final model, suggesting that GPA and retention status were good predictors of future dropouts without considering effects over time.

The model developed by Bowers (2010) influenced the current study in two ways. First, it considered the effects of variables over time. While the final model did not find those effects to be part of an efficient prediction model, that information was helpful to the current study. Second, Bowers' study further revealed two key variables to consider, grades and grade retention, both of which were considered for the current study.

Balfanz et al. (2007) used a data collection and analysis method that most closely aligns with the intentions of the current study. They felt, similarly to the current study, that data collected from surveys or observations can be impractical for schools to use on a wide scale, so they focused on data generally available through school reports. The variables they chose for consideration were standardized test scores, final course grades, end-of-year behavior marks, numbers of in and out-of-school suspensions, attendance, special education status, English as a second language status, and being one or more years overage for grade.

In order to narrow down which variables to use in their prediction model, Balfanz et al. (2007) first applied a two-part test to each variable. They looked for variables that had high predictive power by themselves, which they defined as about 75% or more of students flagged for that variable not graduating from high school. They also looked for variables that had a high yield, which they defined as identifying roughly 10% or more of the future dropouts. Once they determined which variables satisfied both of their tests, they used logistic regression to build their model.

Through their two-part test, Balfanz et al. (2007) identified five variables for further analysis: attendance of 80% or less during sixth grade, a failing grade in math for sixth grade, a failing English grade for sixth grade, receiving an out-of-school suspension during sixth grade, and receiving an unsatisfactory final behavior mark in any subject during sixth grade. They noted specifically that the final behavior mark had an incredibly high yield. Fifty percent of the future dropouts possessed at least one unsatisfactory final behavior mark in sixth grade. Ultimately, Balfanz et al. (2007) dropped suspensions as a variable for their final model since almost all students who were suspended also received an unsatisfactory behavior mark in at least one subject.

Following the two-part test, Balfanz et al. (2007) listed the individual effects of each predictor variable after using logistic regression to control for the effects of the other variables. They reported that students with poor attendance were 68% less likely than other students to graduate, students with poor behavior were 56% less likely to graduate than other students, students who failed math were 54% less likely to graduate, and students who failed English were 42% less likely to graduate. They noted that each variable was a statistically significant predictor of dropout status even after controlling for the other variables.

Combining all variables, Balfanz et al. (2007) noted that their model was able to correctly identify 60% of the students who eventually dropped out of school. In a more descriptive fashion, they discussed the rates of graduation for students with different numbers of predictive flags present. Students with no flags graduated at a rate of 56%. Students with one flag graduated at a rate of 36%. Following that, only 21% of students with two flags graduated, 13% of those with three flags graduated, and only 7% of those with all four flags graduated.

The study by Balfanz et al. (2007) was extremely useful to the current study for a few reasons. First, their study chose variables in a fashion most consistent with the logic used for the current study, namely to focus on practical, school-reported variables. Second, their study focused on data collected during sixth grade. While this is still not as early as the current study intended, it is earlier than the other studies reviewed that used regression analysis. Finally, the reported results from their study gave reason to believe that earlier identification could still be reasonably accurate as compared to later identification.

Conclusion

As was shown in this literature review, there is a need to try to reduce the number of students who drop out before graduation because dropping out leads to increased risks of personal, societal, and familial consequences later in life. History has had evolving viewpoints on how serious the problem of dropouts was, and how to best address it. Many schools, districts, and states developed intervention programs aimed at preventing dropouts, but they did not necessarily have a great impact on the dropout rate. This may have stemmed from several different factors, but one major limitation was that schools did not have tools that could accurately identify potential dropouts. If the intervention programs were not aimed at the right students, they would certainly struggle to be effective.

In addition, there were other limitations to the success of interventions, mainly based on how students were identified. Some identification efforts were too late for programs to be entirely successful. Others were too impractical to apply to all students, thus leaving some students without the necessary interventions. This led researchers to look for earlier, practical, accurate methods for identifying students in need of intervention. Different studies showed varying levels of success, but the basic issue was that earlier identification efforts were generally less accurate than later efforts. With the need, however, to begin interventions earlier, this was still an area very much in need of additional research.

Several of the studies reviewed used logistic regression to analyze their data, which was in line with the intentions of the current study. These studies reported various levels of success at accurately predicting dropouts, but each of them offered insights for the current study. The study by Balfanz et al. (2007) most influenced the current study because of many similarities in philosophy.

All of these factors led to the current study, which was aimed at developing an early, practical, accurate method of identifying potential high school dropouts in elementary school. The literature has provided a great deal of guidance as to which factors to consider and how to develop a research study on the topic. The next chapter will detail the research method for this study.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

The first chapter of this study provided an introduction for the study, including describing many of the problems that tend to occur with increased frequency for students who drop out of school prior to completion of high school. The first chapter also introduced the research questions for the study as well as the conceptual underpinnings guiding the study. It then covered the limitations and assumptions affecting the study, as well as definitions of key terms and the significance of the study. The main focus was on the notion that preventing students from dropping out of school is desirable for many different reasons, and school leaders have felt increased pressure to find ways to decrease the dropout rate.

The second chapter looked more deeply into the background of the dropout issue, including estimates of the problem as well as descriptions of the problems dropouts tend to face in higher frequencies later in life. The chapter then progressed to discussing intervention efforts that schools and districts have started to initiate in an effort to decrease the dropout rate, followed by some explanations for why many of those efforts have not been successful. With the main focus being that interventions were often not started early enough to be successful, the chapter then covered efforts to identify dropouts earlier in school. This included rationale for identifying them earlier as well as general decreases in accuracy the earlier identification was attempted. All of these issues led to the purpose of this study, which was to find ways to identify dropouts earlier while still maintaining a high level of accuracy of identification.

This chapter will present the research questions guiding the study, then proceed to describe the design of the study. This will include the rationales for choices of study design as well as the methods to be employed. The explanation of study design will lead directly to the selection of the population and sample for the study. Rationale will be given for why the population and sample fit the study design, as well as support for why the sample size will suffice for the study.

Following the population and sample selection, data collection will then be covered. This will include how the data will be procured and why the source fits the study well. This section will also include a discussion of the Institutional Review Board (IRB) process and how it applies to this study. The final section of the paper will offer an explanation of how the data will be analyzed. This will include rationale for the choice as well as support through research for the analysis being employed. This study will seek to build an effective predictive model to identify potential dropouts guided by the research questions given in the next section.

Research Questions

For this quantitative study three research questions were used to guide the research:

1. What combinations of school and family variables, such as attendance, grades, test scores, socioeconomic status, discipline, reading level, and grade retention, are the best predictors of the number of high school credits earned?
2. How early are these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) reliable predictors of the number of high school credits earned?

3. How accurately and reliably will a model developed from these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) predict the number of high school credits earned if used as early as fourth grade?

Design for the Study

The approach of this study was a quantitative analysis to address a problem of practice, namely how to accurately predict potential future dropouts at an early age. The study was strictly quantitative because it was intended to develop a model based on data gained from typical school reports. Models that include qualitative elements tend to be less practical for schools, and therefore can keep schools from consistently using the instruments to identify potential dropouts (Balfanz et al., 2007).

The intention was that this study would develop a model that schools could apply to all students every school year to predict the number of high school credits each student would earn. These credit predictions would help identify students at risk of dropping out of school based on whether they would be predicted to earn the required 24 credits to graduate. If a school could typically just run a report that applies to all students, it is hoped that schools would make a consistent practice of working to identify any students at risk of eventually dropping out of school prior to high school graduation. In this way there is less chance that any given student will miss the opportunity to be identified, and therefore the chance to receive needed interventions.

The design of the study was correlational, using multiple regression to build a model to predict the cumulative number of high school credits earned for each student. The state in which this study was located required 24 credits to graduate from high school. As a result, any student predicted to earn less than 24 credits was considered to be at-risk of dropping out of school prior to graduation. Any student predicted to earn 24 or more credits was considered on track to graduate from high school.

Some previous studies have used just simple percentages to tie risk factors to graduation rates (Gleason & Dynarski, 2002; Heppen & Therriault, 2008; Kennelly & Monrad, 2007; Mac Iver, Balfanz, & Byrnes, 2009; Rumberger, 2007), resulting in over-simplified analyses that offered useful information only on a surface level. Other studies have relied on regression for analysis (Balfanz et al., 2007; Bowers, 2010; Suh et al., 2007; Uekawa, Merola, Fernandez, & Porowski, 2010), leading to more in-depth findings. These studies are more closely aligned with the intentions of this study, but they have used data from students later in their school careers than this study.

In order to link the outcome variable (total number of high school credits earned) to the predictor variables from elementary years, it was necessary to collect a longitudinal range of data from elementary years through eventual graduation or dropout. This provided the ability to use elementary data to build a model while already knowing the eventual graduation or dropout status of each student.

Population and Sample

The population for this study was all members of a single graduating year cohort from a suburban Midwest district. The sample included all students for whom data were available from fourth grade (2004) through high school graduation or dropout (2012) within the same district. The rationale for this sample was that graduation or dropout status had to be identifiable, and that information had to be linked to the elementary predictor variables in order to build the predictive model. Students who change schools between elementary and high school years are difficult or impossible to track, and their experiences and treatments differ because they attend different school systems in different areas.

Given the requirement that all students in the sample had available data from elementary through high school years, a large suburban district was chosen in order to present a large possible sample size within one local district. Other studies using regression to build prediction models have used samples as small as 193 (Bowers, 2010) and as large as 41,906 (Uekawa et al., 2010). This study should ideally present a useable sample size of at least 200 students. According to Field (2009), this sample size will be more than sufficient based on several different methods to estimate the needed sample size. He estimated a regression sample would need 10-15 subjects per variable used. With seven independent variables (attendance, grades, test scores, socioeconomic status, discipline, reading level, and grade retention), this would give a minimum sample size of 70-105 subjects. The sample size would still satisfy Field's estimates even if the mobility rate topped 50%, making the assumption of a sufficient sample size reasonable.

Data Collection

This study used archival data from a school district that were gathered through normal educational practices. The data source for this study was a suburban Midwest district. The data were collected for the same students from elementary years (2004) through graduation or dropout (2012) and included identified predictor variables. The rationale for this data source was that this archival data allowed analysis of longitudinal data without requiring a 12 year process of tracking students.

Data Collection Procedures

The data for this study (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) were collected from central office records for the district. Permission was obtained through the superintendent of the district to get access to the records, and collection of these data required working with the core data manager for the district. Compilation of these data was a large undertaking, so the support of the district and the core data manager was crucial. Because the results of this study can potentially be extremely valuable to the district, the work involved in providing the data should have been worthwhile to the district.

Explanation of Variables

This study considered multiple variables as potential predictors of number of high school credits earned. Because each of these variables could have been measured in different ways, an explanation of how each variable was measured is warranted. To that end, following are descriptions of each variable considered for this study.

Attendance. Attendance was used as a percentage of possible school time attended during a given school year. For each student the amount of time attended was divided by the amount of total time school was in session for that year to develop an attendance percentage (Suh et al., 2007). This was treated as a continuous interval variable because values could range anywhere from 0% to 100%.

Grades. Grades were measured as a final grade point average in core classes. Core classes included mathematics, science, social studies, or English classes (Balfanz et al., 2007). Grade point average was calculated on a four point scale, with an A earning four points, a B earning three points, a C earning two points, a D earning one point, and an F earning zero points. All core grades were then averaged to develop a value between 0.0 and 4.0 for each student. This was treated as a continuous interval variable.

Test Scores. Test scores were reported as achievement scores from state-mandated standardized achievement tests. Students took a grade level standardized state test during the fourth, sixth, seventh, and eighth grades in this study. The scores could have ranged from 472 to 849. Test scores, then, were used as continuous interval variables.

Socioeconomic Status. Socioeconomic status was measured by whether or not a student qualified for free, reduced, or full-priced lunch during a given school year (Montes & Lehmann, 2004). Free or reduced lunch is a federally defined status and is not affected by local districts. Free lunch and reduced lunch were combined for reporting purposes, so students were reported as either full price lunch or free/reduced lunch. By combining the two categories this was used as a binary nominal variable, with a value of yes meaning a student qualified for either free or reduced lunch and a value of no meaning a student did not qualify for either. This variable does have one limitation in that students or families must apply for this status, so a qualifying family may not have been reported because there was not an application received by the school.

Discipline. Discipline was measured by the number of occurrences of behavior for which a disciplinary or behavioral log entry was recorded for a student, regardless of severity (Balfanz, Bridgeland, Moore, & Fox, 2010). This included all behaviors logged at the office level. This variable was used as a continuous interval variable reported on a discrete scale. Student numbers of referrals could have ranged from zero to an indefinite number, reported as integers.

Reading Level. Reading level was measured by the district's tool for determining at which grade level a student was reading at the end of each grade level. The district in this study reported reading levels as final reading lexile scores recorded at the end of each grade. Lexile scores could range from zero for a beginning reader to more than 1600 for an advanced reader. Reading scores were used as a continuous interval variable.

Grade Retention. Grade retention was measured by whether or not a student ever had to repeat and complete an entire grade in school prior to fourth grade (Kennelly & Monrad, 2007). A student who ever had to repeat a grade prior to the fourth grade data collection was reported as a yes, while a student who did not have to repeat any grades was reported as a no. Grade retention was used as a binary nominal variable.

Each of these measurements was pulled from the student information system for the district. The system tracked these data points for each student for each school year used in the study. Grades, attendance, and reading level would have been tracked by classroom teachers and reported through the student information system. Test scores, socioeconomic status, discipline, and grade retention would have been tracked by the school office and reported through the student information system. Table 1 lists each of the variables for the study.

Human Subjects Protection

The data for this study were not collected directly from individuals because they were collected from school archival data by the permission of the superintendent. The collected data pertained to living individuals and were considered private. Whether the study was subject to IRB review, then, was determined by whether or not the data were considered individually identifiable. There was no reason that student names needed to be

Table 1

Dependent and Independent Variables

Variable	Dependent/Independent	Type
Total Number of High School Credits Earned	Dependent	Continuous Interval
Attendance	Independent	Continuous Interval
Discipline	Independent	Continuous Interval
Grade Retention	Independent	Binary Nominal
Grades	Independent	Continuous Interval
Reading Level	Independent	Continuous Interval
Socioeconomic Status	Independent	Binary Nominal
Test Scores	Independent	Continuous Interval

tied to the data. It sufficed to assign students a non-identifiable number, such as a Missouri Student Information System (MOSIS) number.

This assignment of a number rather than student names was done to help facilitate IRB approval. IRB approval was sought and granted as exempt based on the specifics of this study. In addition to IRB exemption, the confidentiality and protection of the data were ensured by keeping the data on a secure computer that was kept private.

Data Analysis

As stated previously, the data for this study were pulled from central office core data records. The data were in the form of numerical data for most variables as outlined below and a yes or no response for others. This enabled the data to be fairly simple in form for the amount of data required for this study.

Because the analysis was entirely quantitative, the data were exported from the student information system to Microsoft Excel and collected in spreadsheet form. To analyze the data, the Microsoft Excel spreadsheets were imported into the SPSS (Version 17.0) statistical analysis program. The analysis used predictor variables to develop a multiple regression model that predicted future numbers of cumulative high school credits earned for each student.

The dependent variable for this study was the number of high school credits earned, which could be used to identify whether or not a student qualified to graduate from high school, based on the requirement of 24 credits to graduate. The predictor variables served as the independent variables and were identified from other research studies on this topic. Based on these other studies, the independent variables for this study were attendance, discipline, grade retention, grades, test scores, socioeconomic status, and reading level (Annie E. Casey Foundation, 2010; Balfanz et al., 2007; Bowers, 2010; Carpenter & Ramirez, 2007; Mac Iver et al., 2009; Neild et al., 2007; Suh et al., 2007; Zvoch, 2006).

These variables were used to perform a regression analysis to answer the first research question of this study. Forward stepwise regression was used to identify which variables contributed to the prediction models with significance. After each grade level model was developed, the results were analyzed to identify patterns regarding which variables were retained with significance for each model.

Forward stepwise regression was chosen for the study for one main reason: the development of the prediction models was exploratory in nature. Field (2009) noted that stepwise methods of regression might not be as desirable as other methods because they eliminate the opportunity for the researcher to make methodological decisions about which variables to include in the model being developed. As noted previously, the literature offered minimal guidance about which variables were strong predictors at the elementary level, and studies that did account for elementary predictions did not offer input into the order of significance of any variables. Given these factors, the development of prediction models for this study was exploratory, making forward stepwise regression a reasonable choice.

The second research question of this study dealt with how early the predictors would be valid. Since archival data were collected dating back to elementary school, it was possible to consider data from different ages. Previous research showed that earlier identification attempts tended to result in reduced accuracy (Bowers, 2010). Through the data analysis, it was determined at which grade level the data contributed to the most useful model considering the balance between accuracy and early identification, with the initial goal being to develop the prediction model using data from fourth grade.

The third research question for this study sought to determine how accurate and useful the resulting model would be when developed from fourth grade data. Once it was determined at which elementary grade the best model could be developed, the question of accuracy was answered by analyzing the model developed from fourth grade data. The fourth grade model was analyzed to see how accurately it could predict the number of high school credits earned as well as to see how much variance in credits was explained by the model. Table 2 lists each research question along with the type of analysis used.

To verify validity of the developed prediction model, the model was applied to the existing data. This ensured the model accurately identified future dropouts without missing some or falsely labeling others. The reliability of the model was ensured because the model incorporated report-generated predictor data rather than more subjective data types like surveys or interviews. Because of this, the model yielded the same results each time it was applied to a given set of data. The results of the study should have been generalizable to other suburban Midwest districts of similar demographics. Because the study did not account for data from other types of districts, it was not necessarily generalizable to schools in other settings.

Summary

Dropping out of school has been shown to have many potential consequences for the students involved (Alexander, Entwisle, & Horsey, 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh, Suh, & Houston, 2007). As a result, school officials have put much effort into dropout prevention. Because it is often too late to prevent dropouts once students reach high school, attention has turned toward earlier identification of students at risk of dropping out of school. In this way, schools hope to design interventions to help the students before they ever reach the levels of wanting to drop out of school.

In order to successfully predict students at risk of dropping out, many models have been developed to use as identification tools for schools. Unfortunately, most of the existing tools are lacking in some fashion. This leads to the problem addressed by this

Table 2

Research Questions and Analysis Methods

Research Question	Type of Analysis
1. What combinations of school and family variables, such as attendance, grades, test scores, socioeconomic status, discipline, reading level, and grade retention, are the best predictors of the number of high school credits earned?	Forward stepwise regression performed on data from each grade level from fourth grade through eighth grade, with analysis of the resulting prediction models looking for patterns
2. How early are these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) reliable predictors of the number of high school credits earned?	Forward stepwise regression performed on data from each grade level from fourth grade through eighth grade, with comparisons of the resulting prediction models
3. How accurately and reliably will a model developed from these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) predict the number of high school credits earned if used as early as fourth grade?	Forward stepwise regression performed on the fourth grade data set

study. This study is intended to fill that gap by developing a model that is early, accurate, and practical for schools to use.

The research questions for this study were presented, followed by the description of a quantitative study designed to answer those questions through development of a predictive model. The population and sample selections were presented, along with rationale for why they were appropriate for this study. Data collection procedures were then discussed. For data analysis, a description of the methods planned was presented. This study used multiple regression to build the predictive models. Rationale was given for the selection of this method as well as the adequacy of the planned sample.

The research design and methods in this study were developed to answer the research questions. Through pursuit of this study, a practical predictive model was developed with the intention of offering school leaders a tool they could use to reduce the numbers of dropouts. This could potentially help school officials to accomplish the further goal of alleviating the many problems that may occur when students drop out of school before graduating.

CHAPTER FOUR

RESULTS AND FINDINGS

School leaders are under pressure to keep students in school through graduation. By doing so, those leaders can help students avoid the increased risks of negative characteristics later in life associated with dropping out of high school. These include lower income, need for welfare support, unemployment, and criminal activity (Alexander et al., 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al, 2007).

Those working with students they feel may drop out of school have attempted to identify the students most in need of attention. Christenson and Thurlow (2004) suggested interventions should be aimed at the specific reasons students are at risk of dropping out, showing that such efforts lead to some success at keeping those students in school. One researcher, Bowers (2010), found that interventions applied to students already in high school are not very effective. He felt students at that age were already too far down the path of dropping out of school to effectively change their fates. Gleason and Dynarski (2002) felt that a common problem was targeting the interventions at the wrong students in the first place. Certainly, interventions applied to the wrong students would have little effect on dropout rates.

As a result of these types of findings, school leaders have sought ways to identify potential dropouts both earlier and more accurately. These efforts are based on the findings of researchers like Entwisle and Alexander (1993), who discussed the importance of early educational experiences and their impact on future schooling for students. Suh et al. (2007) noted that earlier interventions are more successful than later interventions, but Bowers (2010) found earlier identifiers of potential dropouts are not as accurate as later identifiers. These last two factors work against each other – the need for earlier identification coupled with the reduction of accuracy that comes with earlier identification. This makes the job of school leaders attempting to intervene even more difficult.

Many researchers have sought ways to more successfully identify potential dropouts earlier (Alexander et al., 1997; Battin-Pearson et al., 2000; Bowers, 2010; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). The current study has attempted to build on the work of these and other researchers to develop a model based on elementary data to identify potential dropouts in need of interventions. If successful, these earlier identifications could lead to more powerful interventions aimed at helping these students persist in school through graduation.

The first three chapters of this study covered the background and purpose of this study, followed by a review of literature pertaining to this research, concluded with a description of the study planned to develop a prediction model. This chapter will discuss the results and findings of the study that was conducted. The data were collected and analyzed, and this chapter will detail the findings of that analysis. It will present an overview of the study, list the research questions that guided the study, and discuss the demographics of the study. Following that, the findings resulting from each of the three research questions will be covered.

Overview of Study

This study was based on a sample of 222 students who entered high school together in a cohort on track to graduate in May 2012 from the same Midwestern suburban high school. The dependent variable in this study was the final number of high school credit hours attained by each student. The independent variables used as predictors for this study were all collected from archived school district data dating back to when the students were in fourth grade in the same district.

The independent variables for the study were identified by reviewing prior research, with a requirement that the data be available for collection through school records rather than through interviews or surveys. The final list of independent variables considered for analysis included attendance, grades, test scores, socioeconomic status, discipline, reading level, and grade retention. Each of these variables was exported from school records into a Microsoft Excel spreadsheet, which was then imported into SPSS for analysis.

Research Questions

For this quantitative study three questions were used to guide the research:

1. What combinations of school and family variables, such as attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention, are the best predictors of the number of high school credits earned?
2. How early are these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) reliable predictors of the number of high school credits earned?
3. How accurately and reliably will a model developed from these variables (attendance, grades, test scores, reading level, socioeconomic status, discipline, and grade retention) predict the number of high school credits earned if used as early as fourth grade?

Demographics

As noted earlier, the sample for this study consisted of 222 students who entered high school as part of the same cohort. This sample was limited to only students who had been in the same district continuously from fourth grade through graduation or dropout so the archived data would be available for all students in the sample. Descriptive analyses were first performed on the data to provide a general overview of the data. While the sample consisted of 222 students, not all variables resulted in 222 values since the archived data were not complete for all students in the sample.

The dependent variable in this study was total number of high school credit hours earned. For the school district from which the data were collected the minimum number of credits required to graduate from high school was 24. For the entire sample of 222 students, 199 students earned 24 or more credits. This represented 89.64% of the total sample. See Figure 1 for a histogram representing frequencies of each number of credits earned in the sample.

Two variables in the study were treated as binary nominal variables. These two variables were socioeconomic status and grade retention. For socioeconomic status, students were listed as having qualified for poverty assistance or not as measured by their participation in the free or reduced lunch price program. For grade retention, students were recorded as having been retained prior to fourth grade or not.

These two variables were only reported one time for all students in the study. As a result there is not a separate summary of these two variables for each of the grade levels considered. A descriptive analysis for all of the other variables in the study will be provided in the following sections, but the frequency analysis of socioeconomic status

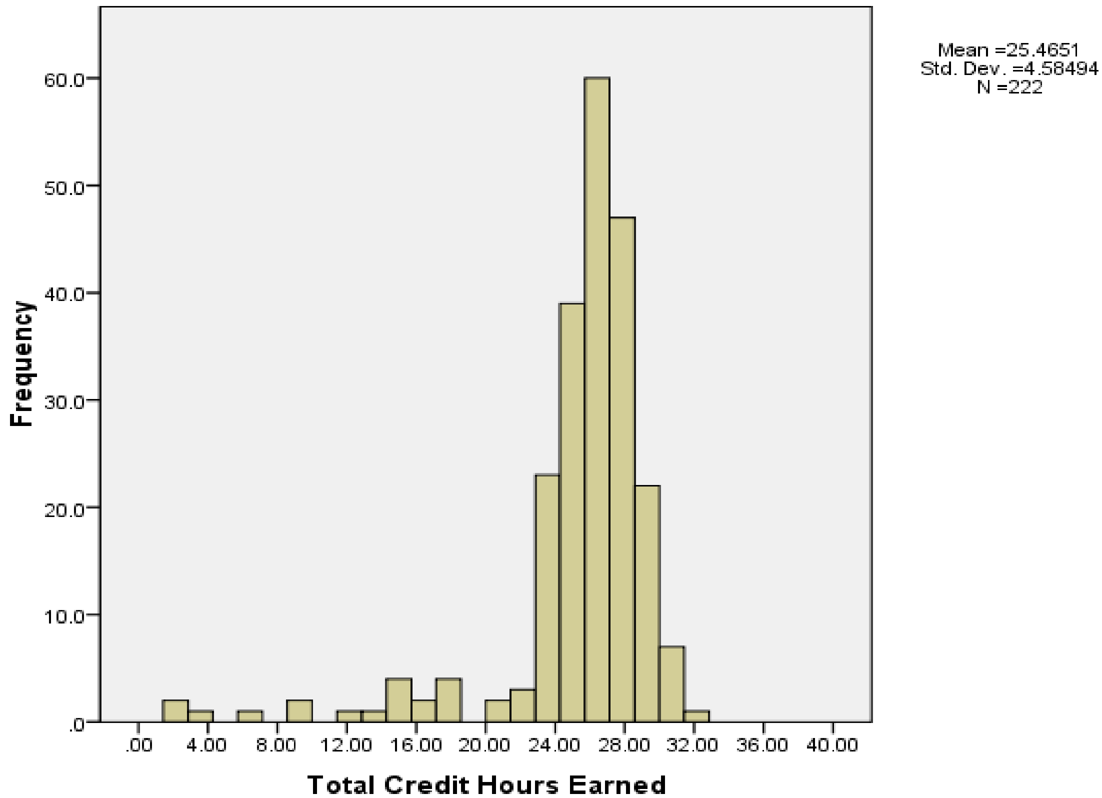

Figure 1. Histogram displaying the total number of credit hours earned by each student and the frequency of each total. Minimum number of credit hours required for graduation is 24.

and grade retention were only performed once for all included grades. See Table 3 for a list of the frequencies of these two variables for the data set.

All other variables in the study were subjected to a descriptive analysis for each grade level of data collected. This includes test scores, discipline, reading level, attendance, and grades. In the following sections, descriptive data for each grade level will be presented, which includes fourth through eighth grades for the graduating cohort.

Table 3

Frequencies of Binary Nominal Variables for all Grades

Variable	*N*	Frequency	Percent
Socioeconomic Status	222	Poverty = 49	Poverty = 22.07%
		Non-Poverty = 173	Non-Poverty = 77.93%
Grade Retention	222	Retained = 3	Retained = 1.35%
		Not Retained = 219	Not Retained = 98.65%

Fourth Grade Data

Of the five continuous variables originally considered, relatively complete fourth grade data were available for four of the variables. One variable in particular, grades, was considerably lacking in the available records. See Table 4 for a listing of the descriptive

Table 4

Descriptives of Continuous Independent Variables (Fourth Grade)

Variable	*N*	Mean	*sd*
Test Scores	205	657.190	27.995
Discipline	222	0.450	1.132
Reading Level	219	837.770	228.680
Attendance	222	0.964	0.030
Grades	67	2.963	0.867

analyses of the continuous variables for fourth grade in this study. This includes test scores, discipline, reading level, attendance, and grades. The lack of complete data on grades will be addressed in a later section, and it will be shown that the prediction model developed from fourth grade data was not severely impacted by the lack of complete grade data due to high correlations with other data.

Fifth Grade Data

Of the five continuous variables originally considered, fifth grade data were available for only three of the variables. The district was not able to recover grade data from that year, and the students did not take any standardized tests during fifth grade, so there were no test scores available. See Table 5 for a listing of the descriptive analyses of the continuous variables for fifth grade data in this study. This includes discipline, reading level, and attendance.

Table 5

Descriptives of Continuous Independent Variables (Fifth Grade)

Variable	*N*	Mean	*sd*
Discipline	221	0.330	0.850
Reading Level	220	69.410	16.318
Attendance	221	0.964	0.034

The lack of test scores for fifth grade was inevitable, since the state in which the study was conducted did not require standardized tests at the fifth grade level at the time the students were in fifth grade. The lack of attendance data was a limitation introduced into the study by the fact the district could not recover any records for that data. The lack of data on grades will be addressed in a later section, and it will be shown that grades tended to show a high correlation with other variables throughout the grade levels in this study, lessening the impact on the prediction model of these missing data.

Sixth Grade Data

Of the five continuous variables originally considered, sixth grade data were available for all five of the variables. See Table 6 for a descriptive analysis of the continuous variables for sixth grade data in this study. This includes test scores, discipline, reading level, attendance, and grades.

Table 6

Descriptives of Continuous Independent Variables (Sixth Grade)

Variable	*N*	Mean	*sd*
Test Scores	219	681.980	26.643
Discipline	219	1.080	2.645
Reading Level	214	963.210	214.613
Attendance	219	0.961	0.044
Grades	219	2.442	1.117

Seventh Grade Data

Of the five continuous variables originally considered, seventh grade data were available for four of the variables. The variables for which data were collected were test scores, discipline, reading level, and grades. For the seventh grade data set, attendance data were not available. This was due to a problem with the district being able to recover the data. The lack of attendance data for seventh grade was a limitation introduced into the study. This limitation will be discussed in more detail in a later section. This limitation likely had an impact on the prediction capabilities of the seventh grade model because attendance was retained as a significant factor in the prediction models developed at each of the other grades that had attendance data. See Table 7 for a descriptive analysis of the continuous variables for seventh grade data in this study. This includes test scores, discipline, reading level, and grades.

Table 7

Descriptives of Continuous Independent Variables (Seventh Grade)

Variable	*N*	Mean	*sd*
Test Scores	216	687.510	29.910
Discipline	216	1.340	3.346
Reading Level	215	1010.850	216.712
Grades	216	2.789	1.053

Eighth Grade Data

Of the five continuous variables originally considered, eighth grade data were available for all five of the variables. See Table 8 for a descriptive analysis of the continuous variables for eighth grade data in this study. This includes test scores, discipline, reading level, attendance, and grades.

Research Question One

The first research question for this study addressed identifying the best combinations of independent variables for predicting total number of high school credits earned. To answer this research question, forward stepwise regression was performed for

Table 8

Descriptives of Continuous Independent Variables (Eighth Grade)

Variable	*N*	Mean	*sd*
Test Scores	218	706.750	28.178
Discipline	219	0.740	1.786
Reading Level	215	1069.270	195.883
Attendance	219	0.959	0.056
Grades	217	2.853	1.005

each grade level data set, from fourth grade through eighth grade. The results of the regression analyses were examined to see which independent variables were retained in each model, and any patterns that emerged. In the following sections, the forward stepwise regression results will be discussed for each grade level of data, followed by a consideration of any patterns that emerged.

Fourth Grade Data Regression Analysis

Multiple regression was performed on the fourth grade data set using the forward stepwise method on the six independent variables that had large enough sample sizes to be considered for the model with the dependent variable of final number of high school credits earned. These included test scores, discipline, reading level, attendance, grade retention, and socioeconomic status. The exclusion of grade data based on too small a sample size will be further discussed in a later section.

Forward stepwise regression was chosen to identify which independent variables contributed to the final model with significance. The final model developed through regression resulted in four variables contributing to the model with significance and two variables being excluded. The final model demonstrated significance in explaining the variance of number of high school credits earned (R=.508, R^2=.258, R_{adj}=.243, F=17.214, p<.001, S_{est}=3.889). The final model accounted for 25.8% of the variance in number of high school credits earned (see Table 9).

Table 9

Forward stepwise regression Model Summary (Fourth Grade)

Model	R	R^2	Adj. R^2	Std. Error of The Estimate	F	p
1	.386	.149	.144	4.134	35.111	<.001
2	.442	.195	.187	4.029	24.259	<.001
3	.477	.227	.216	3.958	19.508	<.001
4	.508	.258	.243	3.889	17.214	<.001

Note. Model 1 entered the variable test scores. Model 2 added the variable attendance. Model 3 added the variable discipline. Model 4 added the variable socioeconomic status.

When predicting total number of high school credits earned at an alpha level of .05, the final model retained four of the independent variables with significance. The included variables were test scores (*Beta* = .261, p < .001), attendance (*Beta* = .207, p = .001), discipline (*Beta* = -.187, p = .003), and socioeconomic status (*Beta* = -.186, p = .005). At the alpha level of .05, two variables were excluded from the model. The two excluded variables were grade retention (*Beta* = .074, *p* = .236) and reading level (*Beta* = .074, *p* = .391). Table 10 summarizes the variables that were included and excluded from the model.

Table 10

Forward stepwise regression Model Variable Results (Fourth Grade)

Variables	*B*	Std Error	Beta	*t*	*p*
(Constant)	-33.336	11.301		-2.950	.004
Test Scores	0.042	0.011	.261	3.925	<.001
Attendance	33.206	9.894	.207	3.356	.001
Discipline	-0.747	0.250	-.187	-2.994	.003
Socioeconomic Status	-2.044	0.714	-.186	-2.865	.005
Reading Level			.074	1.188	.236
Grade Retention			.074	0.860	.391

The final model developed through forward stepwise regression applied to the fourth grade data resulted in the following equation for predicting total number of high school credits earned: Total Credits = .042(Test Scores) + 33.206(Attendance) - .747(Discipline) – 2.044(Socioeconomic Status) – 33.336. The overall fit of this model was R^2 = .258 and the standard error was 3.889. This model was developed without data for grades being included, and the final model excluded the variables grade retention and reading level.

Fifth Grade Data Regression Analysis

Multiple regression was performed on the fifth grade data set using the forward stepwise method on the five independent variables for which fifth grade data were available with the dependent variable of final number of high school credits earned. These included discipline, reading level, attendance, grade retention, and socioeconomic status. The limitations presented by the lack of grade or test score data will be discussed further in a later section.

Forward stepwise regression was chosen to identify which independent variables contributed to the final model with significance. The final model developed through regression resulted in five variables contributing to the model with significance and no variables being excluded. The final model demonstrated significance in explaining the variance of number of high school credits earned (R=.597, R^2=.356, R_{adj}=.341, F=23.695, p<.001, S_{est}=3.689). The final model accounted for 35.6% of the variance in number of high school credits earned (see Table 11).

When predicting total number of high school credits earned at an alpha level of .05, the final model retained five of the independent variables with significance. The included variables were reading level (*Beta* = .293, p < .001), discipline (*Beta* = -.295, p < .001), attendance (*Beta* = .240, p < .001), socioeconomic status (*Beta* = -.224, p < .001), and grade retention (*Beta* = .128, p = .022). At the alpha level of .05, no variables were excluded from the model. Table 12 summarizes the variables that were included in the model.

The final model developed through forward stepwise regression applied to the fifth grade data resulted in the following equation for predicting total number of high

Table 11

Forward stepwise regression Model Summary (Fifth Grade)

Model	R	R^2	Adj. R^2	Std. Error of The Estimate	F	p
1	.387	.149	.146	4.202	38.309	<.001
2	.492	.242	.235	3.976	34.588	<.001
3	.547	.299	.289	3.833	30.662	<.001
4	.583	.340	.328	3.726	27.742	<.001
5	.597	.356	.341	3.689	23.695	<.001

Note. Model 1 entered the variable reading level. Model 2 added the variable discipline. Model 3 added the variable attendance. Model 4 added the variable socioeconomic status. Model 5 added the variable grade retention.

school credits earned: Total Credits = .082(Reading Level) – 1.858(Discipline) + 32.859(Attendance) – 2.457(Socioeconomic Status) + 5.016(Grade Retention) – 10.826. The overall fit of this model was R^2 = .356 and the standard error was 3.689. This model was developed without data for grades or test scores being included, and the final model excluded none of the variables for which fifth grade data was available.

Sixth Grade Data Regression Analysis

Multiple regression was performed on the sixth grade data set using the forward stepwise method on the seven independent variables with the dependent variable of final number of high school credits earned. These included grades, test scores, discipline,

Table 12

Forward stepwise regression Model Variable Results (Fifth Grade)

Variables	*B*	Std Error	Beta	*t*	*p*
(Constant)	-10.826	7.318		-1.479	.141
Reading Level	0.082	0.016	.293	5.075	<.001
Discipline	-1.858	0.350	-.295	-5.307	<.001
Attendance	32.747	7.511	.240	4.375	<.001
Socioeconomic Status	-2.457	0.627	-.224	-3.920	<.001
Grade Retention	5.016	2.181	.128	2.300	.022

reading level, attendance, grade retention, and socioeconomic status. Sixth grade was the earliest grade level in this study for which data was available for all variables.

Forward stepwise regression was chosen to identify which independent variables contributed to the final model with significance. The final model developed through regression resulted in five variables contributing to the model with significance and two variables being excluded. The final model demonstrated significance in explaining the variance of number of high school credits earned (R=.611, R^2=.374, R_{adj}=.359, F=24.822, p<.001, S_{est}=3.388). The final model accounted for 37.4% of the variance in number of high school credits earned (see Table 13).

When predicting total number of high school credits earned at an alpha level of .05, the final model retained five of the independent variables with significance. The

Table 13

Forward stepwise regression Model Summary (Sixth Grade)

Model	*R*	R^2	Adj. R^2	Std. Error of The Estimate	*F*	*p*
1	.519	.269	.265	3.625	77.987	<.001
2	.557	.310	.303	3.531	47.371	<.001
3	.583	.340	.330	3.461	36.040	<.001
4	.602	.362	.350	3.411	29.628	<.001
5	.611	.374	.359	3.388	24.822	<.001

Note. Model 1 entered the variable grades. Model 2 added the variable attendance. Model 3 added the variable reading level. Model 4 added the variable discipline. Model 5 added the variable test scores.

included variables were grades (*Beta* = .313, $p < .001$), attendance (*Beta* = .192, $p = .002$), reading level (*Beta* = .342, $p < .001$), discipline (*Beta* = -.141, $p = .025$), and test scores (*Beta* = -.179, $p = .049$). At the alpha level of .05, two variables were excluded from the model. The two excluded variables were socioeconomic status (*Beta* = -.087, $p = .148$) and grade retention (*Beta* = .086, $p = .139$). Table 14 summarizes the variables that were included and excluded from the model.

The final model developed through forward stepwise regression applied to the sixth grade data resulted in the following equation for predicting total number of high school credits earned: Total Credits = 1.196(Grades) + 19.423(Attendance) + .007(Reading Level) – .244(Discipline) - .029(Test Scores) + 17.823. The overall fit

Table 14

Forward stepwise regression Model Variable Results (Sixth Grade)

Variables	*B*	Std Error	Beta	*t*	*p*
(Constant)	17.823	10.415		1.711	.089
Grades	1.196	0.292	.313	4.101	<.001
Attendance	19.423	6.075	.192	3.197	.002
Reading Level	0.007	0.002	.342	3.723	<.001
Discipline	-0.244	0.108	-.141	-2.252	.025
Test Scores	-0.029	0.015	-.179	-1.984	.049
Socioeconomic Status			-.087	-1.453	.148
Grade Retention			.086	1.485	.139

of this model was R^2 = .374 and the standard error was 3.388. The final model excluded the variables grade retention and reading level.

Seventh Grade Data Regression Analysis

Multiple regression was performed on the seventh grade data set using the forward stepwise method on the six independent variables for which seventh grade data was available with the dependent variable of final number of high school credits earned. These included grades, discipline, reading level, grade retention, and socioeconomic status. The limitations presented by the lack of attendance data will be discussed further in a later section.

Forward stepwise regression was chosen to identify which independent variables contributed to the final model with significance. The final model developed through regression resulted in four variables contributing to the model with significance and two variables being excluded. The final model demonstrated significance in explaining the variance of number of high school credits earned (R=.621, R^2=.386, R_{adj}=.374, F=33.017, p<.001, S_{est}=3.633). The final model accounted for 38.6% of the variance in number of high school credits earned (see Table 15).

Table 15

Forward stepwise regression Model Summary (Seventh Grade)

Model	R	R^2	Adj. R^2	Std. Error of The Estimate	F	p
1	.511	.261	.258	3.958	75.225	<.001
2	.593	.352	.346	3.715	57.564	<.001
3	.612	.375	.366	3.658	42.143	<.001
4	.621	.386	.374	3.633	33.017	<.001

Note. Model 1 entered the variable grades. Model 2 added the variable discipline. Model 3 added the variable test scores. Model 4 added the variable grade retention.

When predicting total number of high school credits earned at an alpha level of .05, the final model retained four of the independent variables with significance. The included variables were grades (*Beta* = .266, p < .001), discipline (*Beta* = -.347, p < .001), test scores (*Beta* = .195, p = .003), and grade retention (*Beta* = .110, p = .050). At the alpha level of .05, two variables were excluded from the model. The two excluded variables were reading level (*Beta* =.076, p = .430) and socioeconomic status (*Beta* = -.081, p = .174). Table 16 summarizes the variables that were included and excluded from the model.

Table 16

Forward stepwise regression Model Variable Results (Seventh Grade)

Variables	*B*	Std Error	Beta	*t*	*p*
(Constant)	2.102	6.347		0.327	.744
Grades	1.161	0.307	.266	3.780	<.001
Discipline	-.475	0.084	-.347	-5.634	<.001
Test Scores	0.030	0.010	.195	3.053	.003
Grade Retention	4.283	2.168	.110	1.975	.050
Reading Level			.076	0.791	.430
Socioeconomic Status			-.081	-1.365	.174

The final model developed through forward stepwise regression applied to the seventh grade data resulted in the following equation for predicting total number of high school credits earned: Total Credits = 2.161(Grades) –.475(Discipline) +.03(Test Scores) + 4.283(Grade Retention) + 2.102. The overall fit of this model was R^2 = .386 and the standard error was 3.633. The final model excluded the variables reading level and socioeconomic status.

Eighth Grade Data Regression Analysis

Multiple regression was performed on the eighth grade data set using the forward stepwise method on the seven independent variables with the dependent variable of final number of high school credits earned. These included grades, test scores, discipline, reading level, attendance, grade retention, and socioeconomic status. Eighth grade was the latest grade level in this study for which data were collected because the focus of the study was to develop prediction models earlier than high school age.

Forward stepwise regression was chosen to identify which independent variables contributed to the final model with significance. The final model developed through regression resulted in four variables contributing to the model with significance and three variables being excluded. The final model demonstrated significance in explaining the variance of number of high school credits earned (R=.634, R^2=.402, R_{adj}=.391, F=35.364, p<.001, S_{est}=3.122). The final model accounted for 40.2% of the variance in number of high school credits earned (see Table 17).

When predicting total number of high school credits earned at an alpha level of .05, the final model retained four of the independent variables with significance. The included variables were grades (*Beta* = .297, p < .001), attendance (*Beta* = .266, p < .001), test scores (*Beta* = .317, p < .001), and grade retention (*Beta* = .121, p = .027). At the alpha level of .05, three variables were excluded from the model. The three excluded variables were reading level (*Beta* = .012, p = .874), discipline (*Beta* = -.111, p = .052), and socioeconomic status (*Beta* = -.084, p = .147). Table 18 summarizes the variables that were included and excluded from the model.

Table 17

Forward stepwise regression Model Summary (Eighth Grade)

Model	*R*	R^2	Adj. R^2	Std. Error of The Estimate	*F*	*p*
1	.514	.264	.261	3.440	76.462	<.001
2	.566	.321	.314	3.314	50.011	<.001
3	.623	.388	.380	3.151	44.670	<.001
4	.634	.402	.391	3.122	35.364	<.001

Note. Model 1 entered the variable grades. Model 2 added the variable attendance. Model 3 added the variable test scores. Model 4 added the variable grade retention.

The final model developed through forward stepwise regression applied to the eighth grade data resulted in the following equation for predicting total number of high school credits earned: Total Credits = 1.179(Grades) + 28.104(Attendance) + .045(Test Scores) + 4.11(Grade Retention) - 36.509. The overall fit of this model was R^2 = .402 and the standard error was 3.122. The final model excluded the variables reading level, discipline, and socioeconomic status.

Research Question One Discussion

To answer the question of which combinations of variables are the best predictors of total number of high school credits earned, forward stepwise regression was performed on the data collected from fourth through eighth grades. Following the regression analysis, the results were examined to identify which variables showed the best predictive

Table 18

Forward stepwise regression Model Variable Results (Eighth Grade)

Variables	*B*	Std Error	Beta	*t*	*p*
(Constant)	-36.509	8.469		-4.311	<.001
Grades	1.179	0.258	.297	4.568	<.001
Attendance	28.104	5.897	.266	4.765	<.001
Test Scores	0.045	0.009	.317	5.066	<.001
Grade Retention	4.110	1.849	.121	2.223	.027
Reading Level			.012	0.159	.874
Discipline			-.111	-1.957	.052
Socioeconomic Status			-.084	-1.454	.147

power. This section will address the patterns that emerged through this analysis of the results.

It is first important to note that only two of the five grade levels examined included data for all seven of the originally considered independent variables. The sixth and eighth grade data sets had values for all variables. The fourth grade set was missing complete data for the grades variable. Fifth grade was lacking both grades and test scores and seventh grade was missing attendance data. This missing data presented new limitations for the study since each of the missing data categories were retained in other grade levels as significant predictors of total number of high school credits earned.

To discuss which combinations of variables were the best predictors of total number of high school credits earned, the resulting prediction models at each grade level were compared to find similarities and differences. Three of the variables were retained in the final prediction models for all grade levels for which the data were available. Test scores were retained in four of the prediction models, and only the fifth grade model was missing test scores data since the students did not take any standardized tests during their fifth grade year. Attendance was retained as a significant predictor in four of the models as well, only missing from the seventh grade model since attendance data were not available from the district for that year. Grades were a significant predictor for three of the models, but were missing from both the fourth and fifth grade models since the data were not available for those grade levels.

The other four variables in the study were available for all grade levels. Discipline was retained as a significant predictor for four of the grade levels and only excluded from the final prediction model for the eighth grade data set. Grade retention was a significant predictor for three of the models, but excluded for both the fourth and sixth grade data sets. Poverty was retained in only two of the models, and excluded from the sixth, seventh, and eighth grade models. Reading level was also only a significant predictor for two grade levels, and was excluded from the final prediction models for the fourth, seventh, and eighth grade data sets. See Table 19 for a summary of the variables retained in each model.

Table 19

Summary of Variables Retained in Each Regression Model

	4th	5th	6th	7th	8th
Grades	N/A	N/A	X	X	X
Attendance	X	X	X	N/A	X
Test Scores	X	N/A	X	X	X
Grade Retention		X		X	X
Reading Level		X	X		
Discipline	X	X	X	X	
Socioeconomic Status		X	X		

Note. N/A in a cell represents data that were not available. Empty cells represent variables that were not retained in the model for that grade level.

From this analysis it was determined that four of the variables presented the best combination of variables for predicting total number of high school credits earned across all grade levels for which data were analyzed. Test scores, attendance, grades, and discipline were present in the most prediction models, and discipline was the only one of those four variables that was excluded from any prediction model. All other variables were excluded more often from final prediction models.

Research Question Two

The second research question for this study asked how early the group of independent variables can be reliable predictors of number of high school credits earned. In order to answer this question, forward stepwise regression was performed on each set of data from fourth through eighth grades. The resulting prediction model for each grade level was examined to determine how strong the prediction model was for each grade level.

As was expected from the literature (Bowers, 2010), the prediction model was strongest when built from data collected during eighth grade and weakest for the fourth grade data. The model explained 40.2% of the variance in number of high school credits earned when built with eighth grade data, 38.6% of the variance when developed from seventh grade data, 37.4% when using sixth grade data, 35.6% when based on fifth grade data, and 25.8% of the variance when the prediction model was based on fourth grade data.

In addition, all of the models had standard errors of at least 3.122. By using two standard errors above and below the predictions to create an interval of values with a prediction accuracy of 95%, this gave a range of total credits earned that spanned greater than 12 credits. Because the number of credits required to graduate was only 24, requiring an interval of 12 credits to gain 95% accuracy was a high standard error.

While none of the prediction models accounted for a large amount of the variance in number of high school credits earned, each model did show significance at predicting credits earned. The question, then, becomes at what level of predictive power a model is considered reliable for educational leaders to use. Because the purpose of the current study was to predict final numbers of credits as early as possible, high school data were not considered.

In comparing the variance in high school credits earned explained by each of the five prediction models, the decrease in explained variance was relatively small from eighth grade to seventh grade (1.6%), from seventh grade to sixth grade (1.2%), and from sixth grade to fifth grade (1.8%). The reduction in explained variance, however, from fifth grade to fourth grade was much larger (10.2%). Given this larger gap, it was determined that the predictive model that struck the best balance between being the earliest prediction with the highest explained variance was the fifth grade prediction model. The explained variance for the fifth grade model was only 4.6% less than that for the eighth grade prediction model while being developed from data for students three years younger. See Table 20 for a summary of the prediction models.

Table 20

Summary of Prediction Models

	8^{th}	7^{th}	6^{th}	5^{th}	4^{th}
Percentage of Explained Variance in the Number of High School Credits Earned	40.2%	38.6%	37.4%	35.6%	25.8%
Standard Error	3.122	3.633	3.388	3.689	3.889

The fifth grade prediction model explained 35.6% of the variance in number of high school credits earned while still giving school leaders a model to apply to elementary-aged students. In addition, for the current study, this model was built with data that were missing two key variables: grades and test scores. As stated earlier, these two variables were retained in all prediction models with significance for all grade levels for which data were available. It would be possible that a fifth grade predictive model could be even stronger if complete data were available.

Even with the low explained variance, the model developed from fifth grade data still provided some practical results. When the model was applied to the fifth grade data set, two useful results surfaced. First, of the 11 students predicted to earn 20 or less credits (24 credits were required to graduate), 8 of them eventually dropped out of school. That means even though the model only accounted for 35.6% of the variance in high school credits earned, it still correctly identified 72.7% of students predicted to earn a low number of credits as dropouts. Second, of the 73 students predicted to earn 27 or more credits, all 73 of them graduated. This means the model also offered practical conclusions for students predicted to earn high numbers of credits. These two results offered some practical guidance for school leaders trying to identify which students are in need of interventions to help them persist to graduation. This also gave more support for the selection of the fifth grade model as the model in this study that offered the best combination of accuracy and early identification.

Research Question Three

The third and final research question for this study addressed how accurately a model developed from fourth grade data could predict total number of high school credits earned. To answer this question the fourth grade data were subjected to the same forward stepwise regression analysis as the other grade levels. Because fourth grade data were the specific targets of one of the research questions, however, more analysis was done. One area in which the fourth grade data were unique in this study was the nature of its missing variable. Data for grades were not complete for the fourth grade data set, but unlike missing data from the other grade levels, partial data were collected for this variable.

Because partial data were available for this variable, it enabled further analysis before just eliminating it from consideration. The low sample size of 67 for the grades variable was of concern for inclusion in the regression model since the minimum estimated sample size needed was 70-105 as noted in Chapter Three. Rather than leaving this variable out of the model without rationale, a correlation was performed between grades and the other continuous independent variables to gauge how closely related grades would be with other variables. The analysis resulted in grades having a correlation of 0.633 with test scores and a correlation of 0.687 with reading level as the two highest correlations (see Table 21).

Table 21

Correlations Between Continuous Independent Variables (Fourth Grade)

	Grades	Test Scores	Discipline	Reading Level	Attendance
Grades	---	.633	-.372	.687	.236
Test Scores		---	-.234	.691	.133
Discipline			---	-.183	-.079
Reading Level				---	.007
Attendance					---

Additionally, a forced entry linear regression was performed with grades entered into the regression model following test scores and again following reading level. When entered following test scores, grades only resulted in an increase in R^2 of 0.004 (from 0.314 to 0.318). When entered following reading level, grades only resulted in an increase in R^2 of 0.025 (from 0.170 to 0.195).

Even though complete data were unavailable for grades, the correlations and regressions were performed in order to better understand the potential effect of this limitation. With a high correlation with two other variables and a minimal increase in the explained variance through regression, it was determined the absence of complete data for grades would not have a large impact on the quality of the model developed through multiple regression. Both test scores and reading level related closely enough to grades to adequately account for the absence of the data in the analysis.

Research question three was answered through forward stepwise regression by determining which variables contributed to the final model with significance at the alpha level of .05 for fourth grade data. As stated earlier in this chapter, four of the independent variables were retained in the model. This included test scores, attendance, discipline, and socioeconomic status. Two of the variables were excluded from the final model. These two variables were reading level and grade retention. The final model using the four variables that were retained demonstrated a significant relationship to the variance in high school credits earned ($p < .001$). The amount of variance in high school credits earned explained by the model was 25.8% ($R^2 = .258$).

Additionally, the final prediction model developed from fourth grade data had a standard error of 3.889. Adding two standard errors above and below the prediction would give a 95% accuracy rate in predicting high school credits earned. This means to predict the total number of high school credits with 95% accuracy this model required a range of roughly 15 credits. Because only 24 credits were required for graduation, a model needing an interval of 15 credits to provide an accurate prediction was not a very strong model. Adding only one standard error above and below the prediction would give a 68% accuracy rate in predicting high school credits earned. Even at this decreased accuracy level, it would still require a range of 7.5 credits to predict the number of credits earned. Again, with only 24 credits required for graduation, a span of 7.5 credits is a large prediction range to accomplish only 68% accuracy.

Summary

As stated in the introduction to this chapter, school leaders and researchers have been attempting to find ways to identify potential dropouts both earlier and more accurately. Much research has considered prediction models applied at the high school level (Battin-Pearson et al., 2000; Gleason & Dynarski, 2002; Heppen and Therriault, 2008; McKee et al., 1998; Rumberger, 2007). A fewer number of studies have developed prediction models at the middle school level (Balfanz et al., 2007; Rumberger, 2007). Very few studies have attempted to predict potential dropouts at the elementary level (Montes and Lehmann, 2004).

This study sought to develop a prediction model to apply to elementary-aged students that was accurate and reliable at identifying future dropouts. In addition to being early, the model was also developed to be practical. To accomplish this, the model only considered data generally contained in school records. This reduced the number of variables being considered for the model to seven.

Research question one was answered by considering data from each of the grade levels from fourth grade through eighth grade. Of the seven variables originally considered, four of the variables seemed to show a consistent significant contribution to the prediction models developed. The four variables that appeared most consistently in the final prediction models were test scores, attendance, grades, and discipline. While each of the prediction models ultimately retained different combinations of variables, these were the four variables chosen as the best combination for predicting the number of high school credits earned.

Research question two was answered by examining the pattern of the amount of variance explained by each model. As expected from the work of Bowers (2010), the amount of explained variance decreased as the age of the students decreased. Eighth grade data yielded the highest explained variance, with a decrease for each grade level model down to fourth grade. The decrease in the amount of explained variance was quite small from grade to grade except for a sizable decrease from fifth grade to fourth grade. As a result, the fifth grade model was chosen as the best predictor due to its balance of two considerations: the explained variance was relatively close to that of the eighth grade data, which had the highest explained variance, and the fifth grade model presented the earliest prediction model that did not sacrifice a large amount of explained variance.

Research question three was answered by more closely examining the fourth grade data set. As noted earlier, forward stepwise regression produced a model that explained only 25.8% of the variance in high school credits earned. This number was not only lower than the other grades considered, it was notably smaller than even the fifth grade model's explained variance. This resulted in a model that did not realistically provide practical guidance for school leaders looking for a strong predictor of high school credits earned based on data collected as early as fourth grade.

The first three chapters of this study presented the background, purpose, review of literature, and research method for the study, guided by three research questions. The fourth chapter has presented an analysis of the data collected, again guided by the three research questions of the study. In addition, answers were provided for each research question based on the analysis of the data. Chapter Five will provide a summary and conclusions for the entire study, and will also include implications for school leaders wishing to reduce school dropout rates and recommendations for future areas to be studied.

CHAPTER FIVE

DISCUSSION

School leaders put a lot of time and effort into keeping students in school through graduation. The main reason for this is that school leaders are under pressure to help students avoid the increased risks of negative characteristics later in life associated with dropping out of high school. These characteristics include lower income, need for welfare support, unemployment, and criminal activity (Alexander et al., 1997; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al, 2007).

School leaders have started by trying to identify which students are most at risk of dropping out of school and therefore are in the most need of attention. Researchers have discussed the various difficulties these leaders have with correctly identifying students in need of help and deciding how best to help them avoid dropping out of school. Christenson and Thurlow (2004) suggested interventions should be aimed at the specific reasons students are at risk of dropping out, showing that such efforts lead to some success at keeping those students in school. One researcher, Bowers (2010), found that interventions applied to students already in high school are not very effective. He felt students at that age were already too far down the path of dropping out of school to effectively change their fates. Gleason and Dynarski (2002) felt that a common problem was targeting the interventions at the wrong students in the first place. Certainly, interventions applied to the wrong students would have little effect on dropout rates.

As a result of these types of findings, school leaders have sought ways to identify potential dropouts both earlier and more accurately. These efforts are based on the findings of researchers like Entwisle and Alexander (1993), who discussed the importance of early educational experiences and their impact on future schooling for students. Suh et al. (2007) noted that earlier interventions are more successful than later interventions, but Bowers (2010) found that earlier identifiers of potential dropouts are not as accurate as later identifiers. These last two factors work against each other – the need for earlier identification coupled with the reduction of accuracy that comes with earlier identification. This makes the job of school leaders attempting to intervene even more difficult.

Many researchers have sought ways to more successfully identify potential dropouts earlier (Alexander et al., 1997; Battin-Pearson et al., 2000; Bowers, 2010; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). This study has attempted to build on the work of these and other researchers to develop a model based on elementary data to identify potential dropouts in need of interventions. If successful, these earlier identifications could lead to more powerful interventions aimed at helping these students persist in school through graduation.

The first four chapters of this study have laid out the background of the study, the guiding research questions, a review of existing literature, the design methodology, and the data analysis. This chapter will conclude, beginning with a summary of the data analysis and a discussion of the results. Following that will be limitations of the current study, implications for practice, and recommendations for future research.

Conclusions

The main goal of this study was to explore a method for predicting future high school dropouts earlier than typical while maintaining an acceptable level of accuracy in identifying these students. In addition, the study attempted to develop this prediction model from readily available school data in an effort to make the model practical. If the model were to be practical it could be of more widespread use to school leaders than methods that are more time consuming. The following section will summarize the results of each of the research questions that guided this study.

Research Question One

The first research question for this study dealt with identifying which combinations of variables would contribute to the most effective prediction model. The seven variables considered for this study were attendance, grades, test scores, discipline, socioeconomic status, reading level, and grade retention. Because data were examined spanning fourth through eighth grades, this question was answered by looking at which variables most consistently contributed to the final prediction models with significance.

Data were not available for all variables at all grade levels, so the results were examined to see if any patterns emerged from the various prediction models based on the available data. Four of the variables seemed to contribute to the final models with significance most often when the data was available: test scores, attendance, grades, and discipline. Considered another way, the other three variables (reading level, socioeconomic status, and grade retention) were more often excluded from the final models than test scores, attendance, grades, and discipline. Research question one was ultimately answered by listing test scores, attendance, grades, and discipline as the combination of variables that created the best prediction model for total number of high school credits earned.

Research Question Two

The second research question for this study examined how early the identified variables created an accurate prediction model for total number of high school credits earned. The main purpose of the entire study was to attempt to develop a prediction model that could be used earlier than current strategies to identify students at risk of dropping out of school. The work of Bowers (2010) created an expectation that earlier prediction models would likely be less accurate than models developed later. This pattern is exactly what the models developed for this study revealed.

Prediction models were developed for each grade level from fourth through eighth grades through multiple regression. Each final model was examined to identify the percentage of variance in high school credits earned explained by the model. To answer the second research question the explained variances of the models were compared. The highest percentage of explained variance was accomplished with the eighth grade model (40.2%). In comparing the different models a pattern emerged when moving from eighth grade toward fourth grade. The seventh grade model explained only 1.6% less variance than the eighth grade model. The sixth grade model showed a decrease in explained variance of 1.2% from the seventh grade model. The fifth grade model explained 1.8% less variance than the sixth grade model. Each of these decreases in explained variance was relatively small. The decrease from fifth grade to fourth grade, however, was considerably larger. The fourth grade model explained 10.2% less variance than the fifth grade model.

Examined another way, the fifth grade model explained only 4.6% less variance in high school credits earned than the eighth grade model, but the decrease from fifth grade to fourth grade was more than double that amount. This led to the final answer for the second research question in this study. In an attempt to balance accuracy of the model with the earliest possible identification, the fifth grade model was chosen as the earliest prediction model that still provided an acceptable level of accuracy as compared to models developed from the data collected when students were older.

Research Question Three

Since the goal of the study was to create the earliest model possible that still provided an accurate prediction, the third research question focused specifically on fourth grade data. Fourth grade data were the earliest data collected for this study, so it was necessary to examine the fourth grade prediction model more closely. As the second research question concluded, the fifth grade prediction model was chosen as a more desirable model than the fourth grade model due to the large decrease in explained variance between the fifth and fourth grade models.

Through multiple regression, a model was built from the available fourth grade data for predicting the total number of high school credits earned. Upon examining the predictive strength of the model, it was determined that the model only explained 25.8% of the variance in high school credits earned. This was a large decrease in explained variance even when compared to the fifth grade model. In addition, it provided an overall level of explained variance that would not realistically provide school leaders with information reliable enough to use when identifying at-risk students.

It should be noted, however, that the fourth grade data set was missing complete data for one variable – grades. This was worth noting because grades was a variable that was included with significance for every model for which the data were available. This could have considerably affected the strength of the model. This limitation will be discussed further in a later section.

Discussion

With the goal of identifying how early a model could accurately predict the number of high school credits earned, this study examined data from one cohort of students who entered high school together. Data were collected from school records for the group of students, with the data spanning all grades from the fourth through eighth grades. The concern that researchers such as Bowers (2010) presented regarding the decreased ability to predict high school dropouts at earlier ages seemed to be confirmed in this study.

The prediction models developed through forward stepwise regression for each grade level showed the expected pattern of a decreased ability to explain the variance in the number of high school credits earned. In addition, even the highest level of explained variance (40.2% in eighth grade) did not offer an accurate enough model to simply hand to school leaders for everyday use. Each of the models would certainly offer, however, a starting point for school administrators as they strive to identify those students in need of extra attention.

While this study attempted to build an acceptable prediction model based on fourth grade data, it did not provide an accurate enough model at that level to be realistic for school leaders. Even though it did not accomplish that specific goal, the study did identify a relatively close alternative. The fifth grade model built from the data was reasonably accurate as compared to the eighth grade data, and still offered a model built from elementary data.

Another positive aspect of the study was that it did build the prediction models from easily retrievable school data, making the model practical for school administrators to use in identifying at-risk students (Balfanz et al., 2007). The study aimed to produce a model that could be applied to an entire grade level of students to identify those at risk as opposed to having to predetermine which students to subject to a more complicated model. While the level of prediction accuracy could have been higher, the results could still offer school leaders a starting place to identify students in need of more analysis rather than choosing those students based on hunches or relying on teacher reports of concerns.

This study did not ultimately produce a model realistically usable by school leaders wishing to identify at-risk students in fourth grade, but it did provide some useful feedback. The techniques used to develop the models differed from other studies in both the variables selected and the method of prediction (Alexander et al., 1997; Battin-Pearson et al., 2000; Bowers, 2010; Christenson & Thurlow, 2004; Gleason & Dynarski, 2002; Suh et al., 2007). By creating models that predicted total high school credits instead of the binary result of graduate versus dropout, the models produced were more sophisticated and capable of taking advantage of the full range of data available.

In addition, the method of creating the models could be easily adapted to any district and any identified variables. The goal was to create a single prediction model at an early grade level, but a specifically developed model could be applied to any number of grade levels to continue to identify students at risk of dropping out of school. This aspect gives the method a more effective ability to be useable and practical for school administrators interested in identifying potential school dropouts (Balfanz et al., 2007).

Limitations

Two main limitations of this study were discussed in Chapter One. One limitation was the inclusion in the study of only those students who had been in the district from fourth grade through graduation or dropping out of school. This eliminated any students who had moved in or out of the district during that time frame. This was a considerable limitation, especially since mobility itself has been shown to be a risk factor for dropping out of school (Jimerson, Anderson, & Whipple, 2002; Montes & Lehmann, 2004; Rumberger & Ah Lim, 2008; Suh et al., 2007).

The second limitation concerned whether the results of the study could be generalized to other settings. While this is a realistic limitation, in the last section it was discussed how this model could be easily developed to be specific to any district and any chosen variables. The ability to easily create new models could help to reduce the effect of this limitation.

Besides the two initial limitations, several other limitations surfaced when the data were collected from the district in this study. The new limitations that were introduced entailed the inability to collect complete data for all grade levels. One variable for which data were not available was test scores at the fifth grade level. These data were not available because at the time the students were in fifth grade, standardized tests were not required, so no test scores existed to report. This limitation could have reduced the ability of the fifth grade model to explain the variance in high school credits earned, especially because test scores were retained with significance for all other grades levels. Had test scores been available in the fifth grade data set, the prediction model might have been more accurate.

Two variables (attendance for 7^{th} grade and grades for 5^{th} grade) were unavailable only because the data were not found. In contrast to the fifth grade test scores, these data were expected to be available, but were not recovered. Both of these missing sets of data might have hurt the respective prediction models. Attendance and grades were each retained with significance for all models for which data were available for the two variables.

The final limitation that surfaced when collecting data was the inability to collect complete grades data for the fourth grade data set. Unlike the other limitations, these data were partially available, but not in a large enough sample size to be usable in the regression model. The presence of partial data enabled further analysis in an effort to see how much the lack of complete data might hurt the predictive ability of the model. This analysis seemed to show that the lack of the complete data on grades might have been accounted for with other variables, but one fact remains: grades were retained with significance for all models for which complete data were available. This might have caused the fourth grade prediction model to be less effective than it could have been otherwise.

Implications for Practice

The results of this study offer several implications for practice. Currently, the literature is lacking in dropout prediction models applicable to elementary students. The goal of this study was to produce a model that could accurately predict high school dropouts as early as elementary school. While practitioners may desire a model with a stronger predictive power, this study did create a model that could be applied at any grade level. The study also succeeded in developing a model intended to be practical for school leaders.

The implications that result from these two factors would be based on how school administrators use the findings. With the lack of current studies identifying elementary students at-risk of eventually dropping out of school, the models developed through this study provide something more concrete for school leaders to use than more subjective methods. Even though models built from earlier grade levels might show a decreased ability to accurately predict future school dropouts, the models could be used as a starting point for further analysis of students.

The practicality of the models offers potential help in this area. Because the model could be easily applied to an entire grade level of students, even a model with lower accuracy at predicting dropouts could identify some students on whom leaders could focus. Instead of relying on reports of teacher concerns or other observations, school officials could use this type of model as an initial indicator of risk.

A related implication of this study is based on the fact that this study actually predicted number of high school credits earned as an indicator of graduate status, as opposed to the simple binary variable of graduate versus dropout. The value in this difference impacts the degree with which school leaders could use the model even when the explained variance is not incredibly high. Even if the model may not reliably predict high school credits earned within a narrow range, for students who are predicted to earn an extremely low number of credits, the model might hold more ability to identify which students might be at risk of dropping out of school later in their educational careers.

To illustrate this, the model developed from the fifth grade data, Total Credits = .082(Reading Level) – 1.858(Discipline) + 32.859(Attendance) – 2.457(Socioeconomic Status) + 5.016(Grade Retention) – 10.826, was applied to the fifth grade data set. When the values of the variables were entered into the formula and the resulting predicted numbers of credits were sorted, two interesting patterns emerged. First, of the 11 students predicted to earn 20 or less credits (24 credits were required to graduate), 8 of them eventually dropped out of school. Even though the fifth grade model only explained 35.6% of the variance in high school credits earned, when looking for a low threshold of predicted credits earned, the model correctly predicted 72.7% of dropouts. At the other end of the spectrum, of the 73 students predicted to earn at least 27 credits, all 73 of them graduated. Even with a low explained variance, this model offered useful information for school leaders wishing to identify which students to more closely monitor.

This final implication was perhaps the most practical implication to come out of this study. While the techniques used to create the models could be adapted to any school district and any grade level, if the models would not offer usable information the results would not help school leaders. To know that they could at least identify the majority of students in the direst need of help would be a valuable starting point for preventing future dropouts.

In order to for school leaders to benefit from this research, a discussion of how to use the information is warranted. Ideally, a prediction model would be developed specifically for a school district based on their own data. This may differ from models developed for other districts, both in the variables included and the grade level for which the most effective model is chosen. Once a final model is developed, the district leaders would then use the identified model for predicting potential dropouts in subsequent years. In a given building, the model would be applied to all students in the chosen grade. Upon calculating predicted credits for each student based on their own data, decisions about which students need interventions could be made.

In this way, school leaders could look at students in real time. For example, fifth grade students could be identified based on their current data. Once the predicted numbers of high school credits are calculated, school officials could determine which students are at risk of dropping out of school at some point by identifying the students predicted to earn extremely low numbers of credits. They could then target their interventions at the identified students beginning in elementary school. By doing this, school leaders could hope to impact potential dropouts, helping them to get back on track for graduation.

Recommendations for Future Research

The results of this study have led to several recommendations for future research. Many of the recommendations fall in the area of improving the research methods themselves. The first recommendation is based on the fact that this study focused on seven independent variables identified through the literature review as being important predictors of high school dropouts. It is certainly possible that other variables exist that could be stronger predictors of high school dropouts or number of credits earned. Part of the recommendation would be to consider a wider range of variables. This study limited the variables to those easily collected through typical school records in an effort to create a practical model. As a result, many variables were not considered for this study.

Further research could be performed with the same basic model, but including different variables for consideration. One example of a variable that was excluded from consideration in this study was mobility. Mobility was identified by multiple researchers as being a predictor of dropouts (Jimerson, Anderson, & Whipple, 2002; Montes & Lehmann, 2004; Rumberger & Ah Lim, 2008; Suh et al., 2007), yet the methods used in this study did not permit its use as a variable that was analyzed. A study using different methods to collect the data may allow for the inclusion of mobility.

Another recommendation related to improving the research methods of this study would be more complete data collection. Perhaps the most effective way to ensure complete data collection would be to perform the study longitudinally as the students progress through school. This would enable the researcher to collect the necessary data without the inherit risks involved when collecting archived data: that the data were never collected in the first place or that the ability to collect the data became hampered.

A final recommendation for further research related to improving the current study would be to collect data on a wider range of students in different settings. The data in this study came from only one suburban, Midwest district. Further research that considers other districts in different settings would offer insights into the ability to generalize the findings.

Perhaps the most significant of the recommendations for future research lies in the area of what follows the findings of this study. While this study was intended to help school leaders identify potential dropouts at earlier ages, the study never addressed what to do for students identified as being at-risk of dropping out of school. It was shown that earlier interventions are more effective than later interventions at preventing students from dropping out of school, but specific strategies were not discussed in depth. Further research could take a closer look at current intervention strategies.

More importantly, though, future research could consider interventions specific to elementary students. If students at-risk of dropping out of school have typically not been identified as early as elementary school, strategies to intervene would certainly be lacking. Further research into strategies aimed at helping elementary students get back on track for graduation is recommended.

Finally, further research is recommended in the area of targeting interventions. Kennelly and Monrad (2007) discussed intervention efforts targeted at specific deficiencies. Further research could be performed in two areas related to this topic. This would include developing a prediction model that not only predicts dropouts earlier, but also indicates specific areas of concern. It would also include developing interventions aimed at addressing those areas of concern. If both of these ideas were addressed through future research, school leaders would gain some tools to use in the efforts to prevent dropouts.

The research in this study has provided useful information for school administrators as they attempt to prevent dropouts. The results, however, do not provide the complete set of tools necessary to fully implement intervention strategies. Future research in several related areas could provide those tools. Then school leaders could have improved practical resources available to them as they push for lower and lower dropout rates.

References

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.

Annie E. Casey Foundation. (2010). *Early warning! Why reading by the end of third grade matters* (A KIDS COUNT Special Report). Baltimore, MD: Fiester, L.

Balfanz, R., Bridgeland, J. M., Moore, L. A., & Fox, J. H. (2010). *Building a grad nation: Progress and challenge in ending the high school dropout epidemic* (A report by Civic Enterprises and the Everyone Graduates Center at Johns Hopkins University). 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.

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*, 568-582.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, *103*, 191-207.

Caraway, K., Tucker, C. M., Reinke, W. M., & Hall, C. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools*, *40*, 417-427.

Carpenter, D. M., & Ramirez, A. (2007). More than one gap: Dropout rate gaps between and among black, Hispanic, and white students. *Journal of Advanced Academics, 19*, 32-64.

Christenson, S. L., & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science, 13*, 36-39.

Dynarski, M., & Gleason, P. (1998). *How can we help? What we have learned from evaluations of federal dropout-prevention programs* (A Research Report from the School Dropout Demonstration Assistance Program Evaluation). Princeton, NJ: Mathematica Policy Research.

Entwisle, D. R., & Alexander, K. L. (1993). Entry into schools: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology, 19*, 401-423.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Fitzpatrick, K. M., & Yoels, W. C. (1992). Policy, school structure, and sociodemographic effects on statewide high school dropout rates. *Sociology of Education, 65*, 76-93.

Franks, A. (2007). *Introduction to Erik Erikson's stages of psychosocial development*. Retrieved July 31, 2011, from http://www.helium.com/items/543768-introduction-to-erik-eriksons-stages-of-psychosocial-development

Gleason, P., & Dynarski, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Students Placed at Risk, 7*, 25-41.

Hauser, R. M. & Koenig, J. A. (2011). *High school dropout, graduation, and completion rates.* Washington, D.C.: The National Academies Press.

Heppen, J. B., & Therriault, S. B. (2008). *Developing early warning systems to identify potential high school dropouts* (Issue Brief). National High School Center.

Janosz, M., LeBlanc, M., Boulerice, B., & Tremblay, R. E. (1997). Disentangling the weight of school dropout predictors: A test on two longitudinal samples. *Journal of Youth and Adolescence, 26*, 733-762.

Jerald, C. (2006). *Identifying potential dropouts: Key lessons for building an early warning data system*. Washington, D.C.: Achieve.

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*, 441-457.

John W. Gardner Center for Youth and Their Communities. (2011). *Using early warning systems to predict and prevent dropout* (Youth Data Archive Policy Factsheet). Stanford, CA: Stanford University School of Education.

Kennelly, L., & Monrad, M. (2007). *Approaches to dropout prevention: Heeding early warning signs with appropriate interventions* (Report). National High School Center.

Mac Iver, M. A., Balfanz, R., & Byrnes, V. (2009). *Dropouts in the Denver public schools: Early warning signals and possibilities for prevention and recovery* (Report contributing to the Colorado Statewide Dropout Initiative). Baltimore, MD: Johns Hopkins University.

McKee, J. M., Melvin, K. B., Ditoro, V, & McKee, S. P. (1998). SARIS: Student at-risk identification scale. *The Journal of At-Risk Issues*, *4*(2), 24-32.

Miami-Dade County Public Schools. (2007). *Student mobility* (Information Capsule). Miami: Research Services. Retrieved July 31, 2011, from http://drs.dadeschools.net/InfoCapsules/IC0608.pdf

Missouri Department of Elementary and Secondary Education. (2012). *School food services: News and updates.* Retrieved May 16, 2012, from http://dese.mo.gov/divadm/food/

Montes, G., & Lehmann, C. (2004). *Who will drop out from school? Key predictors from the literature* (Technical Report and Works in Progress Series: Number T04-001). Rochester, NY: Children's Institute.

Neild, R. C., Balfanz, R., & Herzog, L. (2007). An early warning system. *Educational Leadership, 65,* 28-33.

Reschly, A. L., & Christenson, S. L. (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. (2007). *Early predictors of high school graduation and dropout* (Statistical Brief 3). Santa Barbara, CA: California Dropout Research Project.

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

Rumberger, R. W., & Ah Lim, S. (2008). *Why students drop out of school: A review of 25 years of research* (Policy Brief 15). Santa Barbara, CA: California Dropout Research Project.

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

Schargel, F. P., & Smink, J. (2001). *Strategies to help solve our school dropout problem.* Larchmont, NY: Eye on Education.

Stearns, E., & Glennie, E. J. (2006). When and why dropouts leave high school. *Youth Society, 38,* 29-57.

Suh, S., Suh, J., & Houston, I. (2007). Predictors of categorical at-risk high school dropouts. *Journal of Counseling & Development, 85*, 196-203.

Uekawa, K., Merola, S., Fernandez, F., & Porowski, A. (2010). *Creating an early warning system: Predictors of dropout in Delaware* (REL Mid-Atlantic Technical Assistance Brief). Delaware: Regional Educational Laboratory Mid-Atlantic.

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*, 91-117.

VITA

Nathan Hoven was born January 1, 1976, in Forest City, Iowa. He moved to Maryland Heights, Missouri, at a young age, and was raised there. After graduating from Pattonville High School in 1994, Nathan attended the University of Missouri – Rolla and received his Bachelor's Degree in Applied Mathematics with certification to teach high school math. He went on to receive his Master's Degree from the University of Missouri – St. Louis in Secondary School Administration. Shortly after college Nathan married his wife, Kelly, and they have three children: Nicholas, Zachary, and Danielle. They currently reside in O'Fallon, Missouri.

Nathan's professional career has always involved public K-12 education. He began as a math teacher at Hazelwood Central High School in Florissant, Missouri. He then moved to teach math at Pattonville High School in Maryland Heights, Missouri, where he also went on to become an assistant principal in the same building. Following that, Nathan moved to Rolla, Missouri, where he served as the principal of Rolla High School. Nathan currently serves as the principal of Timberland High School in Wentzville, Missouri. In addition to time spent teaching math and working as an administrator, Nathan also had experience coaching football, wrestling, and golf at the high school level.

Nathan chose to pursue his EdD in Educational Leadership and Policy Analysis through the University of Missouri - Columbia to further himself as a school administrator. His focus throughout the program has been to find practical applications for his role in K-12 education. He has also given presentations at several conferences in an effort to share what he has learned, and he intends to continue his work at finding ways to improve education for all students.