

# Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts<sup>1</sup>

Alex J. Bowers<sup>2,3</sup>  
bowers@tc.edu

## ABSTRACT

Studies of student risk of school dropout have shown that current predictors of “at-risk” status do not accurately identify a large percentage of students who eventually dropout. Through the analysis of the entire grade 1-12 longitudinal cohort-based grading histories of the class of 2006 for two school districts in the United States, this study extends past longitudinal conceptions of dropout to a longitudinal *risk* perspective, using survival analysis, life tables and discrete-time hazard modeling to appropriately account for student graduation, transfer or dropout. The risk of dropout began in grade 7, with the most hazardous years at grades 8 and 11. A novel calculation of teacher assigned grades, non-cumulative GPA, is identified as a strong predictor of student dropout.

**KEYWORDS:** Dropout, dropout prediction, graduation, at-risk, grades (scholastic), retention, survival analysis, discrete-time hazard modeling, logistic regression, longitudinal studies

## INTRODUCTION

### Overview

Graduation from high school in the United States is known to lead to improved life outcomes for students, as opposed to dropping out of school or obtaining an alternative diploma (Berkthold & Carroll, 1998; Greene & Caire, 2001; Kienzi & Kena, 2006; Tyler, 2003). However, reporting of student graduation rates has recently become a topic of much debate (Greene & Winters, 2005; Heckman & LaFontaine, 2007; NCES, 2004; Sable, Gaviola, & Hoffman, 2007; Swanson, 2004a, 2004b; Viadero, 2006). While the debate continues over estimations of national and state-level graduation rates, school districts are in need of improved systems to help identify and assist students at risk of not graduating on time (Hammond, Linton, Smink, & Drew, 2007; Orfield, 2004) before the act of dropping out of school occurs.

The research to date examining student graduation and dropout rates has focused on both large-scale estimation of national dropout rates, as well as the issue of early prediction of dropouts. For the 2003-2004 school year, the United States Department of Education estimated a national graduation rate of 74.3%

(Seastrom, Hoffman, Chapman, & Stillwell, 2006), and that data is supported by other studies that have also estimated national average graduation rates above 70% (Greene & Caire, 2001; Greene & Winters, 2005). However, other recent studies have begun to reexamine the methods of national graduation estimation and have reported national average graduation rates below 70% (Swanson, 2004a, 2004b). Although estimated averages for states are useful for reporting and policy purposes, examination of individual school district graduation rates and how best to predict failure to graduate has begun to come to the fore in the literature (Balfanz, Herzog, & MacIver, 2007; Balfanz & Legters, 2004; Hammond et al., 2007).

In the research that has attempted to examine which students dropout, there has been a focus on identifying early indicators of potential student dropouts to help schools focus resources for children that may be at risk of dropping out of school (Alexander, Entwisle, & Kabbani, 2001; Allensworth & Easton, 2005; Balfanz et al., 2007; Barrington & Hendricks, 1989; Dynarski & Gleason, 2002; Finn, 1989; Fitzsimmons, Cheever, Leonard, & Macunovich, 1969; Gleason & Dynarski, 2002; Jimerson, Egeland, Sroufe, & Carlson, 2000; Laird, DeBell, & Chapman, 2006; Lloyd, 1974, 1978; Rumberger, 1995). For many of these studies recently, they are situated in large United States urban school districts. For instance, in Chicago, researchers have found that the district graduation rate is 54% and that graduation rates differ by ethnic group in that 39% of African American students graduate, in comparison to 51% of Hispanics, 71% of European Americans and 85% of Asian students (Allensworth, 2005). Additionally, in Chicago, receiving a failing grade in one or more courses as well as having a low number of credits by the end of ninth grade was predictive of a student not graduating and this was especially problematic for males and Hispanic students (Allensworth & Easton, 2005; Roderick & Camburn, 1999). In a related study, representative samples of Baltimore district students were examined over their lifetime within the district’s schools and many longitudinal factors were found that were useful in predicting risk of dropout, including grade retention, low academic achievement, and family socioeconomic status (Alexander et al., 2001). Together, these studies indicate that specific predictors of student graduation do exist for school districts and that these predictors can be identified by ninth grade or earlier. However, while many longitudinal factors are known to influence the probability that a student will become “at-risk” for not graduating on time from high school (Rumberger, 2001, 2004b), most of these factors have been shown to be fairly inefficient and variable predictors of student risk (Gleason & Dynarski, 2002; Hammond et al., 2007). These issues with early dropout identification are especially problematic given recent efforts to design and assess dropout prevention programs (for a review, see Hammond et al., 2007). The purpose of this study is to extend and improve upon the current longitudinal perspective of dropout identification to a longitudinal *risk* perspective through the use of teacher assigned grades to better understand and identify which students will dropout and when.

<sup>1</sup> This is a preprint of this manuscript, published in 2010 in the *Journal of Educational Research*. Recommended citation: Bowers, A.J. (2010) Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *The Journal of Educational Research*, 103(3), 191-207. [doi:10.1080/00220670903382970](https://doi.org/10.1080/00220670903382970)

<sup>2</sup> Teachers College, Columbia University; Bowers@tc.edu; 525 W. 120<sup>th</sup> Street, New York, New York 10027.

ORCID: 0000-0002-5140-6428, ResearcherID: C-1557-2013

<sup>3</sup> Formerly at The University of Texas at San Antonio at time of publication.

Note: This document last updated August 23, 2013

### *Identification of Students At-Risk of Dropping out of School*

Retaining a student at any grade level is one of the highest predictors of dropping out (Abrams & Haney, 2004; Jimerson et al., 2005; Laird et al., 2006; Montes & Lehmann, 2004; Roderick, Nagaoka, Bacon, & Easton, 2000). Yet, other than retention, the literature on risk factors that predict student dropout also includes many other variables that have been tested for the ability to assign students as “at-risk” with the purpose of predicting, and ultimately preventing future student dropouts. For much of the early dropout literature, four main factors predicting student dropout were identified including academic achievement, as measured by teacher assigned grades, absenteeism, retention and family socio-economic status (Barrington & Hendricks, 1989; Ensminger & Slusarcick, 1992; Finn, 1989; Fitzsimmons et al., 1969; Lloyd, 1974, 1978; Rumberger, 1995). These findings have recently been replicated (Alexander et al., 2001; Balfanz et al., 2007). However, across the literature to date, the predictive validity of these risk factors has been shown to be relatively low (Dynarski & Gleason, 2002; Gleason & Dynarski, 2002; Hammond et al., 2007). Additional risk factors also explored are a single parent home, family on public assistance, sibling drop out, absenteeism, disciplinary problems, failing grades at the high school level or middle school level, or overage for grade-level, among others (Allensworth, 2005; Balfanz et al., 2007; Bradley & Lenton, 2007; Eckstein & Wolpin, 1999; Gleason & Dynarski, 2002; Laird et al., 2006; Montes & Lehmann, 2004; Roderick, 1993; Rumberger, 2004b). Nevertheless, for many of these variables, individual dropout rates for students with each risk factor have been shown to be below 10% of the students with that risk factor at the middle school level, and below 30% at the high school level (Gleason & Dynarski, 2002; Hammond et al., 2007; Laird et al., 2006; Montes & Lehmann, 2004; Weber, 1989). Additionally, Gleason and Dynarski (2002) have shown that when many of these factors are combined using multivariate statistics, the percentage of students identified with the multivariate prediction variable who ultimately drop out only rises to 23% at the middle school level, and 42% at the high school level. Thus, many of these risk factors only accurately identify a subset of the students who ultimately dropout.

These studies are limited in that the vast majority of the research to date includes data only on students at the high school level and does not account for the longitudinal nature of the data or the dropout problem. This is problematic. If identification of potential dropouts does not occur until high school, the deleterious impact of these risk factors over the extended period of time before high school is not assessed or included when judging early risk factors. The literature on student’s lack of motivation to stay in school indicates that the decision to dropout is not based on a single factor or moment, but rather is the cumulative effect of multiple risk factors, influencing the student over long periods of time within a district (Alexander et al., 2001; Jimerson et al., 2000; Randolph & Orthner, 2006). For many districts nation-wide, 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. This study argues for an early preventative intervention approach, rather than focusing on students one or two years before they may dropout, or after dropout has occurred.

### *The Use of Teacher Assigned Grades to Predict Dropout Risk*

While these multiple variables for predicting student dropout have been nominated and tested in the literature to date with varying

results, teacher assigned grades have consistently been identified as a useful variable in predicting student dropout (Barrington & Hendricks, 1989; Eckstein & Wolpin, 1999; Finn, 1989; Fitzsimmons et al., 1969; Lloyd, 1974, 1978; Rumberger, 1995). However, the definition of a “grade” has differed from study to study. Overall cumulative grade point average (GPA) has been incorporated into multiple different prediction statistics, and at the high school level, low grades are moderately predictive of student dropout (Gleason & Dynarski, 2002) with 27% of the students with low grades dropping out. Also with a focus on the high school years, receiving a failing grade in any course appears to be associated with a higher likelihood of a student dropping out of school (Allensworth & Easton, 2005). More recently, of students who dropped out, about 20% failed either mathematics or English at the sixth grade level (Balfanz et al., 2007). While these studies examining failing grades at either the high school or middle school level are useful, the argument of this study is that this past research ignores the longitudinal nature of the dataset available and does not use the entire grading scale of teacher assigned grades across multiple courses.

To date, the most systematic examination of dropout prediction and the utility of teacher assigned grades was conducted by Eckstein et. al. (1999). “Grades” were defined as cumulative GPA in five core courses (mathematics, English, social studies, science and foreign language) from grade levels 8 through 12. They showed that grades were highly predictive of dropout, in that in their sample, four types of students existed; students who received on average high grades (A-) throughout high school and graduated on time, students who received average medium grades (C+) throughout high school and graduated on time, students who received lower average grades (C-) and dropped out throughout high school, and students who received low average grades (D+) and dropped out mostly before grade 9. This showed that rather than examine failing grades only, intervals of GPA should be considered, in which low grades (D+ or less) could be predictive of students dropping out of school. Additionally, these results indicated that two high-risk time-periods exist for students who may drop out of school, at eighth grade and mid-way through high school. However, these results are problematic for three main reasons. First, the data was from a 1979 survey of students and has not been repeated with a more recent dataset. Second, the initial research questions posed by Eckstein et. al. focused on the potential link between number of hours worked outside of school by students and dropout, which they found marginal evidence for, rather than a specific focus to identify potential early variables that could identify students at risk of dropping out of school. Third, their model, a structured logistic regression model, was over-parameterized, incorporating 92 parameters, making it difficult to identify individual parameters, such as teacher assigned grades, that, on their own, may be useful for school districts in identifying current students at-risk of dropping out of school.

Thus, these issues in the dropout identification literature underscore the need for more research in this area for five reasons. First, there is a need for analysis of recent datasets, analyzing these same issues with dropout prediction for recently graduated or dropped out students. Second, acknowledging the longitudinal nature of a student’s decision to dropout of school as well as the ability to use the longitudinal data that exists for students in schools, recent innovations in longitudinal data analysis, such as the use in this study of risk analysis and discrete time hazard modeling, can be used to help test which variables are most predictive of a student’s risk of dropping out of school. Third,

building on the work of Eckstein et. al. (1999), there is a need to analyze teacher assigned grades in a systematic fashion, rather than as single course failures or GPA in selected courses only at the high school level, to replicate and extend the findings in a more parsimonious model to identify how predictive teacher assigned grades can be in identifying students who may drop out of school. Overall, the aim of this study is to detail an analysis of the usefulness of teacher assigned grades, from grade 1 through grade 12, in predicting the likelihood of students dropping out of school as a way to provide a single useful variable for school districts to identify not only the students who are most likely to dropout of school, but also at which grade levels those students are most likely to dropout.

## METHOD

For this study, the entire teacher assigned subject-specific grade histories of a sample of students were collected. The sample of students was comprised of all of the students of the entire cohort of the class of 2006 (whether or not they graduated) for two districts, West Oak and South Pine (pseudonyms). Although the overall sample size of students for this study is small ( $n=193$ ), the two districts that agreed to participate in the study were selected based on their willingness to participate and the availability of the data for students who either graduated or dropped out of school<sup>1</sup>. Both districts are located in the United States Mid-West, are located within 10 miles of each other, have contiguous borders with each other and are first ring suburbs of a large metropolitan area. Due to the requirements imposed by both districts to allow access to the student data, and for issues of confidentiality, throughout this study, district specifics, including the  $n$  of subgroups, are intentionally left vague. In addition, as recommended for determining the minimum sample size required for the method used here, survival analysis (Lakatos, 1988), *a priori* power analysis calculations assuming a power of 0.8,  $\alpha = 0.05$ , and an overall predicted rate from the literature of survival to graduation of 75%, indicated a required sample size of 183, which was exceeded.

### *District Context*

West Oak is defined as a mid-sized central city by the U.S. census, with less than 3000 students attending two elementary schools, a middle school and a high school. In 2006, the district served an overall student population that was about 70% economically disadvantaged, 50% Hispanic, 30% European American and 15% African American. The district has historically lagged behind the state averages on state standardized tests in both reading and mathematics at all grade levels (NCES, 2006; S&P, 2006).

South Pine is defined as an urban fringe of a mid-sized city by the U.S. census, with fewer than 3000 students attending three elementary schools, a middle school and a high school. In 2006, the district served a student population that was about 50% economically disadvantaged, 50% European American, 20% Hispanic, and 15% African American. The district has historically scored near the state averages on state standardized tests in both reading and mathematics at all grade levels (NCES, 2006; S&P, 2006).

### *Data Collection:*

Data collected included the entire longitudinal grading histories from grade 1 through grade 12 for all subjects for each student who had ever been on-track to graduate from high school in June of

2006. Students were included in the sample if they started grade 1 on-track to graduate in 2006, whether or not they actually did graduate. For both districts, the grade 1 school year was 1994/1995. This resulted in an  $n$  of 193, 103 from West Oak and 90 from South Pine.

Each student's permanent record in paper form was accessed from the district's long-term paper file storage (i.e. report cards). Student data was entered into SPSS. For each student, grades for every subject for every year were recorded, 1 through 12. Because it was outside of the scope of this study, attendance was not recorded. Letter grades for each subject at each grade level were converted into the following numeric grading scale: A=4.0, A- = 3.666, B+ = 3.333, B = 3.0, B- = 2.666, C+ = 2.333, C = 2.0, C- = 1.666, D+ = 1.333, D = 1.0, D- = 0.666, E or F = 0. Mean non-cumulative grade point averages (GPA) for each grade level were then calculated by calculating the mean GPA for all subjects within each grade level.

Additional variables were also recorded for each student, including gender, ethnicity, if the student had ever been retained, and if the student had graduated on time with their cohort or had dropped out. The issue of the designation of "dropout" is highly contested in the literature (Greene & Winters, 2005; NCES, 2004; Swanson, 2004a, 2004b; Viadero, 2006) and official definitions differ by state and by region. Nevertheless, many students who were on track to graduate on time with their cohort in this sample did not. Because the term "dropout" is currently under contention in the literature and policy domains, for this study, as has been previously recommended (Bowers, 2007; Ensminger & Slusarcick, 1992; Marrow, 1986), dropout designation was handled in the following way. Students were categorized into three mutually exclusive groups; on time graduation, censored, or dropped out. Students were designated as graduating on time if there was a record of a diploma awarded in June of 2006. Students were censored from the dataset if there was a record of transfer to another school district, or if the student was still enrolled in either district at the end of grade 12 but did not graduate (indicating that the student was behind their cohort in credits, but was on-track to graduate from high school in 5 years, rather than 4)<sup>2</sup>. The remaining students were designated as having dropped out of school. Thus, the focus of this study, and the designation of students as "dropout", is aimed at students who stopped attending school with their cohort in either district and thus were unable to graduate on time with a regular high school diploma.

While these three designations may seem fairly straightforward, the options presented to students in the U.S. system who do not wish to graduate on time are many, and exact categorization of dropout or not is difficult due to these multiple options (Swanson, 2004a). These issues were handled in the following manner for this study. A valid student transfer was defined as any student's record which contained a request for student transcripts for student transfer to another school district or school which was not an alternative school. A record of a transcript request from an alternative school was defined as a non-valid indicator of student transfer for on time high school graduation, and thus was an indicator of the educational challenges faced by the student with a high probability that the student would not graduate on time with their cohort in June of 2006. Lacking confirming graduation or alternative degree completion data from the alternative education schools, it cannot be determined if the students who transferred to alternative education programs graduated on time with their cohort with a full high school diploma, rather than a G.E.D. It is the case

that many students who transferred to alternative high schools had low or failing grades in multiple subjects at the time of the transfer. Past research on the G.E.D. option has shown that it is not equivalent to a regular high school diploma (Cameron & Heckman, 1993; Tyler, 2003) and thus is not considered for this study as on time graduation with a standard high school diploma. Even if these students did graduate from an alternative high school with a diploma or an alternative high school degree (G.E.D.), this study is focused on the on-time graduation of the cohort of students in a traditional high school program, and so thus will consider students who transferred to an alternative education program as having dropped out of the regular high school program.

If a student's file did not contain a record of a diploma award, a request for student records from another district, or the record ended prematurely, that student was designated as a student dropout. Thus, this categorization of students dropping out of school may contain some unknown degree of false positives; students who are categorized as dropout but did graduate on time in some other district which had not requested that student's transcripts from the two districts in this study. While the false positive issue is a threat to the internal validity of the conclusions of this study because the number of false positives cannot be estimated, dropout, as defined here, is a reasonable designation given that the majority of the students coded dropout did have records of either non-attendance, refusing to attend school, incarceration or expulsion. In this way, dropout, while not a "pure" indication of students who opted to stop attending school, should be considered a reasonable proxy.

#### *Statistical Analyses:*

Following the methods recommended for longitudinal data analysis (Singer & Willet, 2003), the person-level dataset was converted into a person-period dataset, with event defined as student dropout at the time when the student's academic record ended with either school district. Students who graduated on time were censored at grade 12, while students who had a record of a valid transfer to another school district were censored at the end of the last academic year they were present in either school district. Students who transferred into the district at any time between grades 1 and 12 were considered late entrants, and were included into the study with missing data for all variables other than grade-level and dropout event variables up to the late entry time point. Students who were retained at any grade-level and whose records ended before June of 2006 without requests for transcripts from a different school district were considered dropout, while students who did not graduate on time but were still enrolled in the districts due to retention or future planned graduation after summer school or a fifth year of high school, were censored at the end of their current grade in June of 2006.

As suggested previously to most effectively study time dependent effects on a student's risk of dropping out of school (Bradley & Lenton, 2007; Willett & Singer, 1991), the effects of multiple variables on a student's probability of dropout were estimated employing survival and discrete-time hazard analysis using logistic regression with the person-period dataset (Bradley & Lenton, 2007; Singer & Willet, 1993, 2003; Willett & Singer, 1991, 1995). Time invariant dichotomous variables included in the analysis were: "Gender", coded 0 for females, 1 for males; "Not European American/Ethnicity", coded 0 for European Americans and 1 for all other ethnic groups; and "District", coded 0 for South Pine and 1 for West Oak. Three time variant variables were included. In the

variable "Retained", students retained at any grade level, 1 through 12, were coded as 1 in the time-period in which they were retained and then all other periods thereafter, all other students were coded as 0. The continuous variable "DEtotal" recorded the total number of letter grades "D" or less in all subjects for each time period. The variable "Non-Cumulative GPA" contained the mean student grade point averages for each time period for all subjects, non-cumulative. Logit discrete-time hazard models were estimated and fit to the dropout grade-level data, as well as the calculation of life tables, estimated hazard and survival functions, and median lifetimes using the methods detailed by Singer & Willet (2003).

## RESULTS

### *Descriptive Statistics*

The entire grading and enrollment histories from grades 1 through 12 were recorded for two cohorts of students. These two cohorts were the graduating classes of 2006 from two school districts in the industrial mid-west, West Oak and South Pine (pseudonyms). For the overall sample,  $n=193$  students, which included all of the students who were ever enrolled in either district who could have graduated with their cohort in 2006 (*see methods*); 103 having attended West Oak and 90 having attended South Pine (*Table 1*). The percentages of female and male students differed somewhat between the two cohorts, as did student ethnicity, with the majority of students in the West Oak cohort of Hispanic ethnicity while the majority of students in the South Pine cohort were of European American ethnicity (*Table 1*). Due to the vagaries of district data collection and data retention, while many student records included data such as ethnicity, for both districts, multiple students did not have any ethnicity recorded. This issue with missing ethnicity data was most prevalent for the West Oak cohort, with 31.1% of the student records containing no information on ethnicity.

**Table 1:** *Descriptive variables and frequencies by district*

|                          | West Oak          | South Pine       |
|--------------------------|-------------------|------------------|
| Total Number of Students | 103               | 90               |
| Sampled                  |                   |                  |
| Gender (%)               |                   |                  |
| Female                   | 40.8%<br>(42/103) | 60.0%<br>(54/90) |
| Male                     | 59.2%<br>(61/103) | 40.0%<br>(36/90) |
| Ethnicity (%)            |                   |                  |
| European American        | 29.1%<br>(30/103) | 62.2%<br>(56/90) |
| Hispanic                 | 30.1%<br>(31/103) | 8.9%<br>(8/90)   |
| African American         | 9.7%<br>(10/103)  | 10.0%<br>(9/90)  |
| Asian                    | 0%<br>(0/103)     | 4.4%<br>(4/90)   |
| No Ethnicity Data        | 31.1%<br>(32/103) | 14.4%<br>(13/90) |

### *Student Dropout*

As described in the methods, student dropout was mainly defined as students whose academic records ended prematurely in either district before June of 2006 graduation from high school without a valid record of transfer. For the entire sample, 75.6% of the

students graduated on time with a full high school diploma, while 24.4% of the students dropped out. By district, the on-time graduation rate for West Oak was 65.0%, and for South Pine was 87.8%. This data corresponds to previous research that has shown that graduation rates are highly variable district to district, and falls both above and below the 74.3% graduation rate reported for the United States as a whole (Greene & Caire, 2001; Greene & Winters, 2005; Rumberger, 1995; Seastrom et al., 2006; Swanson, 2004a, 2004b).

Data on multiple variables were recorded for each student in the sample. Table 2 presents descriptive data for the dichotomous variables Gender, Not European American, District and Retained. Data is presented for the overall dataset as well as disaggregated by dropout status. For the overall dataset, nearly half of the students were female and half were male. While both districts had ethnically diverse student cohorts (*see Table 1*), due to the need for dichotomous non-multicategorical variables in the discrete-time hazard model described below, ethnicity was dichotomized for the overall dataset as either European American or all other ethnic groups. For the dataset, 11.4% of the students were retained at some time during their time in either district (*Table 2*).

**Table 2:** Dichotomous variables and percentages of students who graduated on time or were dropouts, by variable

|                           | Overall            | On-Time Graduation | Dropped Out       |
|---------------------------|--------------------|--------------------|-------------------|
| Gender (%)                |                    |                    |                   |
| Female                    | 49.7%<br>(96/193)  | 81.3%<br>(78/96)   | 18.8%<br>(18/96)  |
| Male                      | 50.3%<br>(97/193)  | 70.1%<br>(68/97)   | 29.9%<br>(29/97)  |
| Not European American (%) |                    |                    |                   |
| European American         | 44.6%<br>(86/193)  | 88.4%<br>(76/86)   | 11.6%<br>(10/86)  |
| Not European American     | 32.1%<br>(62/193)  | 79.0%<br>(49/62)   | 30.6%<br>(19/62)  |
| District (%)              |                    |                    |                   |
| West Oak                  | 53.4%<br>(103/193) | 65.0%<br>(67/103)  | 35.0%<br>(36/103) |
| South Pine                | 46.6%<br>(90/193)  | 87.8%<br>(79/90)   | 12.2%<br>(11/90)  |
| Retained (%)              |                    |                    |                   |
| Never Retained            | 88.6%<br>(171/193) | 84.2%<br>(144/171) | 15.8%<br>(27/171) |
| Retained in any Grade     | 11.4%<br>(22/193)  | 9.1%<br>(2/22)     | 90.9%<br>(20/22)  |

Disaggregating the data by on-time graduation or dropout reveals striking differences within each of the categorical variables in Table 2. While female and male students were nearly evenly split in the sample, 29.9% of males dropped out of school in comparison to only 18.8% of females. Similarly, non-European American students disproportionately dropped out in comparison with European American students. District rates of dropout also varied, with West Oak having higher rates of student dropouts than South Pine (*Table 2*). These findings replicate previous studies and extend the past national and large urban district findings to the context of small first-ring suburbs.

The literature to date on dropouts and on-time graduation has indicated that student grade retention is a strong predictor of a student not graduating on time (Jimerson, Anderson, & Whipple, Bowers (2010) *Grades and Graduation*

2002; Jimerson et al., 2005; Laird et al., 2006; Montes & Lehmann, 2004; Roderick, 1993; Roderick & Camburn, 1999; Roderick et al., 2000). For this study, 90.9% of the students who were retained in any grade level dropped out (*Table 2, right bottom*). This result confirms and extends the previously reported negative impact of repeating a grade-level on graduation rates to small urban and first-ring suburban districts (Alexander et al., 2001; Jimerson et al., 2002; Jimerson et al., 2005; Roderick, 1994; Roderick et al., 2000), demonstrating that retention at any level in districts of the size studied here may not be serving students in a way that promotes increased achievement and eventual graduation.

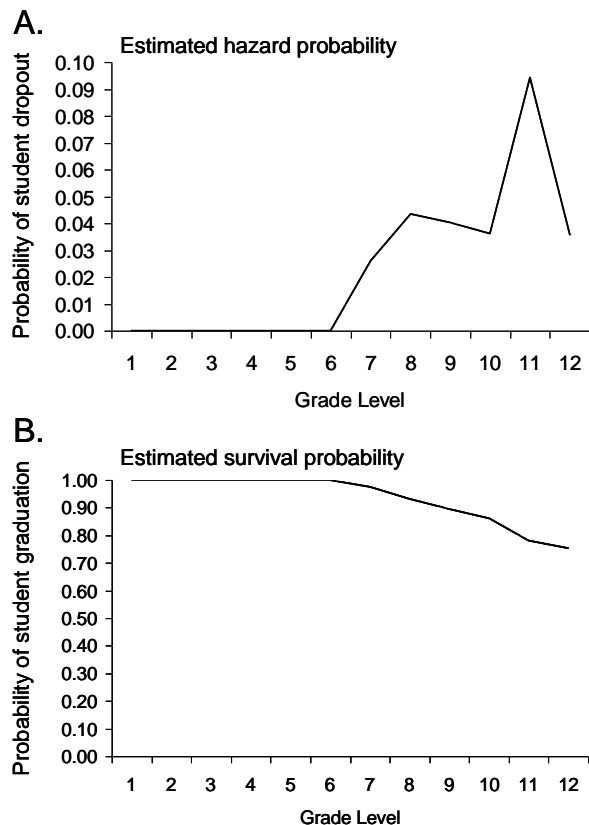
While the data presented here replicates and extends the results of past studies that have shown similar patterns of student dropout in large urban contexts, dropout data presented as aggregated overall rates of dropping out does not acknowledge the time-sensitive nature of the schooling and dropout process. Rather, while past standard practice has been to calculate overall dropout rates as percentages of students who have remained in school versus those who have not over an entire period, such as aggregating grades 9 through 12 as was done in Table 2, better methods have been nominated to deal with dropout data. These methods are able to account for the conditional nature of dropout rates over each year of schooling, as well as more appropriately handling data of students who leave the dataset with unknown outcomes such as transfers to other school districts. It is this issue, and the uses of survival analysis and discrete-time hazard analysis, which is taken up next.

#### *Survival Analysis of Student On-Time Graduation or Dropout*

To investigate the event occurrence of students leaving school early through dropout, survival analysis using a discrete-time hazard model was utilized as has been suggested in the longitudinal data analysis literature (Singer & Willet, 1993, 2003; Willett & Singer, 1991), and more recently in the emerging dropout literature (Bradley & Lenton, 2007). Such analyses have been shown to be superior to simple means and weighted means when analyzing the risk of a terminal event (Singer & Willet, 2003; Willett & Singer, 1991), such as dropping out of school in which a student, once he or she has dropped out, can not reverse their status and become a student who has never dropped out. Survival analysis allows one to examine the students in each grade level still at risk of leaving school and not graduating on time, rather than aggregating together all years in the dataset. This requires the removal of two types of students from calculations at each grade level who have already left the school before that grade level, since they can no longer be at initial risk of an event that has already happened to them. One type is students who dropped out. The other type is students who left the school but had a valid transfer to another school district. These students' risk of dropping out once they left West Oak or South Pine could no longer be determined and so they were censored (removed from the dataset) from the calculations at those grade levels and beyond. As suggested by Singer and Willett (2003), Table 3 presents a life table detailing these dropout event histories for grades 1 through 12 of the sample of 193 students in the dataset. A life table with hazard and survival estimates, as is presented here, is superior to past statistical techniques used to describe the dropout event. This is due to the way in which dropouts and transferred students are described in the life table, as either experiencing the event and then being removed from subsequent time points, or becoming unknown for the event and being censored, which provides a more

**Table 3:** Life table for the event histories for student dropout of the sample of 193 students

| Grade | Number                                 |  |  | Proportion of   |   |
|-------|--|--|--|---|---|
|       | enrolled at the end of the school year | who dropped out during the school year | censored at the end of the school year | students at the beginning of the school year who dropped out before the start of the next school year (Hazard Estimate) | all students still enrolled at the end of the school year (Survival Estimate) |
| 1     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 2     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 3     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 4     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 5     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 6     | 193                                    | 0                                      | 0                                      | 0.0000  | 1.0000  |
| 7     | 186                                    | 5                                      | 2                                      | 0.0262  | 0.9738  |
| 8     | 176                                    | 8                                      | 9                                      | 0.0435  | 0.9315  |
| 9     | 167                                    | 7                                      | 19                                     | 0.0402  | 0.8940  |
| 10    | 160                                    | 6                                      | 27                                     | 0.0361  | 0.8617  |
| 11    | 144                                    | 15                                     | 34                                     | 0.0943  | 0.7804  |
| 12    | 135                                    | 5                                      | 53                                     | 0.0357  | 0.7525  |



**Figure 1:** Estimated hazard and survival functions for the overall dataset. The estimated hazard function shows that the probability of dropout is zero until grade 7, and then rises thereafter, with the most hazardous year at grade 11, at which almost 10% of students are at risk of dropping out (Panel A). The estimated survival function shows a cumulative decline in the probability of graduating on time begins at grade 7, drops steadily over the subsequent grades, and that in grade 12 over 70% of the students do not experience the dropout event, and graduate on time (Panel B).

Bowers (2010) *Grades and Graduation*

realistic and complete numerical description over time than past methods. For grades 1 through 6, no student in the dataset left school. After grade 6, a number of students in each grade level left school and were categorized as dropouts (Table 3, third column). Additionally, beginning in grade 7, students began to transfer to other school districts, and thus their dropout status became unknown and they were censored from the dataset (Table 3, fourth column). Table 3 also presents the estimated hazard and survival probabilities (Table 3, fifth & sixth columns). As suggested for longitudinal data of this type (Singer & Willet, 2003), graphing the estimated hazard and survival probabilities allows for greater ease in examining and interpreting each of these functions (Figure 1).

The estimated hazard probability shows the proportion of students in the sample at each grade level still at risk of dropping out (i.e., all students who were still enrolled at the beginning of that academic school year) who left school within that grade level and were thus categorized as a student dropout (Table 3, fifth column; Figure 1, A). This type of analysis is read as the percentage risk of experiencing the event at each specific time point for the dataset. Additionally, plotting the hazard function allows for the visual identification and interpretation of the trend of students experiencing the event “dropout” across the time periods, visually identifying peak timepoints while controlling for the conditional nature of the event (i.e. students who dropped out in grade 8 are no longer present in the dataset for grade 9). For example, a hazard probability of 0.0361 in the fifth column of Table 3 for grade 10, and graphed in Figure 1A at the grade 10 time point, indicates that for the individuals still in the dataset by grade 10, 3.61% of them experienced the event of dropping out.

Using these life table calculations, estimating the hazard function for dropping out, the probability of dropping out for this dataset is zero until grade 7, at which point it rises to 2.6%. The percentage of students at risk of dropping out continues to increase over the subsequent years, appearing to level off somewhat across grades 8 through 10 at about 4%, peaking dramatically at 9.4% in grade 11, and then falling to 3.5% in grade 12. These data replicate past research indicating that the middle school years are important to consider when examining dropout rates (Balfanz et al., 2007;

Rumberger, 1995). This analysis, for the first time in the U.S. context, also quantifies the hazard and survival estimates as calculations of the percent of students at risk of experiencing the event while controlling for students who have left the dataset through already experiencing the event (already dropping out) or leaving the dataset (censored due to transfer to another district). Risk of dropping out began in middle school at grade 7, with 2.6% of the students dropping out. The percentage then rises over time, with the most hazardous years at grades 8 and 11, 4.4% and 9.4% of students dropping out, respectively. However, if students in this dataset remained in school into grade 12, their risk of dropping out fell to 3.6% (*Table 3, fifth column; Figure 1, A*).

The final far-right column of Table 3, the survival function, presents the data in a cumulative manner, displaying the data points as the percentage of the full sample who survived, i.e., those who did not leave school and did eventually graduate on time, appropriately controlling for students who already dropped out of the dataset (*Table 3, sixth column; Figure 1, B*). At grade 7, 97.4% of the sample was still enrolled in the two districts. This number decreased over time and by the end of grade 9 with 89.4% of the students still enrolled. By grade 12, 75.3% of the students were still enrolled by the end of the school year and graduated on time with their cohort. This number can be considered the graduation rate of these two districts together. As previously suggested (Willett & Singer, 1991), it is argued here that this method of estimating a survival function, from grades 1 through 12 as a graduation rate, is superior to previously articulated methods of calculating graduation rates as end-product statistics (Greene & Winters, 2005; Laird et al., 2006; Seastrom et al., 2006; Swanson, 2004a, 2004b). This is due to the use in survival analysis of student-level data from grades 1 through 12, the ability to easily handle transfer students through censoring, and the more accurate estimates provided by survival analysis over standard descriptive statistical techniques.

While the calculation of overall graduation and dropout rates using survival analysis and life tables with this sample are of interest, these numbers do not give an indication of which students are most at risk of dropping out, only when. The main focus of this study is to combine the above techniques of examining the timing of dropout with a quantification of the usefulness of teacher assigned grades, as a previously identified variable in the dropout literature, in helping to predict which students identified in the above analysis at each time point were most at risk of dropping out. In this way, the aim of this study is to improve upon past dropout identification research and use an appropriate longitudinal analysis technique to analyze if teacher assigned grades are useful to school districts as predictors of students at risk of dropping out as well as identifying the times when students are most at risk of dropping out.

#### *Teacher Assigned Grades and the Prediction of Dropout*

As discussed above, previous research has indicated that low and failing teacher assigned grades may be a useful predictor of student dropout (Alexander et al., 2001; Allensworth, 2005; Allensworth & Easton, 2005; Balfanz et al., 2007; Eckstein & Wolpin, 1999; Roderick, 1993; Roderick & Camburn, 1999). However much of this literature has been focused on general descriptive characteristics of the data, the high school level, grades in specific core subjects, the number of failing grades per grade level, or sub-samples of students from large urban districts. This study addresses these issues in four ways. First, rather than a small sample of students from a much larger district-wide population, such as

previous research which sampled approximately 12% of a student cohort from a large urban district (Alexander et al., 2001), for this study 100% of the students in a cohort from two different school districts were analyzed. Second, past dropout studies have focused on specific grade levels, middle school or high school, or created different statistical models for each of the different levels (Alexander et al., 2001; Balfanz et al., 2007; Gleason & Dynarski, 2002). As will be discussed below, this study analyzes the risk of dropping out using a discrete-time hazard model, which appropriately handles student dropout and transfer data, as presented above, and models and tests the effect of time at each grade-level on the risk of dropping out. Third, the literature that has explored teacher assigned grades as predictors of student dropout has overly focused on specific subjects, such as English and mathematics (Balfanz et al., 2007; Eckstein & Wolpin, 1999), rather than on considering grades in all courses for each grade level.

Fourth, and of most interest to this study, much of the literature on grades and dropouts has overly focused on overall counts of the number of course failures (Allensworth & Easton, 2005; Balfanz et al., 2007), rather than on using the entire traditional 4.0 grading scale data available. Additionally, of the few studies that have examined dropout risk statistically using the 4.0 grading scale, the variable used has been cumulative grade point average (GPA) (Eckstein & Wolpin, 1999; Gleason & Dynarski, 2002; Roderick, 1993). This is problematic due to the discrete time nature of the act of dropping out. Using a cumulative variable for grades over time, such as GPA, masks the inherent ups and downs in the data that could occur within any one year that could indicate that a student is challenged with school. In addition, acknowledging the work that has shown that the middle school years are exceptionally problematic for students (Balfanz et al., 2007; Rumberger, 1995), the use of GPA, traditionally calculated cumulatively for only the high school grades of 9 through 12, does not provide any information on grades in middle school. Thus, it is argued here that GPA, considered as a cumulative variable to date, should rather be calculated non-cumulatively, creating a non-cumulative GPA variable that reflects a student's grades across all subjects within one grade level at a time. Consequently, there is a need for a study on the usefulness of teacher assigned grades for predicting student dropout that addresses each of these four issues together. The present study addresses these issues by examining all students within cohorts (i.e. all potential dropouts, rather than a sample), testing the effect of time on the risk of dropping out, as well as analyzing teacher assigned grades as effective predictors of student dropout, using all subjects, the full grading scale, and non-cumulative GPA.

#### *Discrete Time Hazard Modeling of Student Dropout as a Time Varying Event*

To address the question of the extent to which different variables, including teacher assigned grades, may correspond to a student's increased risk of dropout, discrete-time hazard models using logit regression were utilized. As has been argued previously, a discrete-time hazard model using logit regression is more appropriate for predicting a student's risk of dropout from some set of variables, while controlling for demographic variables (Bradley & Lenton, 2007; Singer & Willet, 2003; Willett & Singer, 1991). This is because of the dichotomous outcome variable, dropout or not, the appropriate handling of transfer students through censoring, and the ability to include the effect of time within the equation – modeling the discrete nature of time, due to grade-level changes,

**Table 4:** Discrete-time hazard model fitting by logistic regression for students at risk of dropout

|  | Model A              | Model B              | Model C              | Model D              | Model E              | Model F              | Model G              |
|--|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Parameter Estimates and Asymptotic Standard Errors |                      |                      |                      |                      |                      |                      |                      |
| Grade 7  | -3.616***<br>(0.453) | -2.620***<br>(0.595) | -2.854***<br>(0.607) | 1.150<br>(1.004)     |                      |                      |                      |
| Grade 8  | -3.091***<br>(0.361) | -2.299***<br>(0.521) | -2.769***<br>(0.538) | 1.426<br>(0.961)     |                      |                      |                      |
| Grade 9  | -3.172***<br>(0.386) | -1.843***<br>(0.435) | -2.954***<br>(0.578) | 2.538*<br>(1.028)    | 0.066<br>(0.665)     |                      |                      |
| Grade 10   | -3.283***<br>(0.416) | -2.034***<br>(0.472) | -3.675***<br>(0.653) | 1.759<br>(1.051)     | -0.362<br>(0.668)    |                      |                      |
| Grade 11   | -2.252***<br>(0.271) | -1.494***<br>(0.385) | -2.385***<br>(0.462) | 2.390**<br>(0.919)   | 1.304**<br>(0.494)   | 1.287**<br>(0.393)   |                      |
| Grade 12   | -3.296***<br>(0.455) | -2.215***<br>(0.523) | -2.831***<br>(0.571) | 2.309*<br>(0.943)    | 0.690<br>(0.618)     |                      |                      |
| Gender   |                      | -2.182***<br>(0.238) | -2.339***<br>(0.259) | -0.552<br>(0.457)    |                      |                      |                      |
| Ethnicity  |                      | -1.812***<br>(0.245) | -1.327***<br>(0.262) | 0.277<br>(0.507)     |                      |                      |                      |
| District   |                      | -1.467***<br>(0.239) | -1.774***<br>(0.274) | 0.825<br>(0.521)     |                      |                      |                      |
| DEtotal  |                      |                      | 0.356***<br>(0.090)  | -0.346*<br>(0.138)   | 0.004<br>(0.075)     |                      |                      |
| Retained   |                      |                      | 3.297***<br>(0.517)  | 3.160***<br>(0.545)  | 2.262***<br>(0.433)  | 2.205***<br>(0.398)  | 2.325***<br>(0.385)  |
| Non Cumulative GPA                                 |                      |                      |                      | -2.615***<br>(0.307) | -1.970***<br>(0.145) | -1.943***<br>(0.123) | -1.826***<br>(0.107) |
| <i>N</i>   | 2172                 | 1768                 | 1493                 | 1493                 | 1655                 | 1655                 | 1655                 |
| n parameters                                       | 6                    | 9                    | 11                   | 12                   | 7                    | 3                    | 2                    |
| -2Log-likelihood                                   | 1970.267             | 794.584              | 684.323              | 158.199              | 250.143              | 252.173              | 261.814              |
| Chi-Square   | 1040.764             | 1653.385             | 1385.415             | 1911.539             | 2044.174             | 2042.144             | 2032.503             |
| Cox & Snell pseudo-R <sup>2</sup>                  | 0.381                | 0.608                | 0.605                | 0.722                | 0.709                | 0.709                | 0.707                |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

within the equation. Rather than experiencing a continuous change over time, students experience school in discrete sections, one grade-level at a time, with open time periods in the summer between each discrete period. Thus, a discrete-time hazard model is appropriate for modeling dropout risk and testing different variables for the ability to predict which students may dropout, specifically testing the extent to which teacher assigned grades are predictive of student dropout.

As noted above, to calculate student risk of dropout, many variables have been nominated in the literature and shown to be marginally effective (Allensworth, 2005; Gleason & Dynarski, 2002; Hammond et al., 2007; Laird et al., 2006; Montes & Lehmann, 2004). For this study, multiple variables were analyzed for their ability to estimate student risk of dropout over time, including time, as modeled by the effect of each grade level from grade 7 through 12, Gender, Ethnicity, District, DEtotal, Retained, and Non-cumulative GPA (*see methods*). For these last three variables, DEtotal is a simple count of the total number of D, or E/F letter grades for each student within each school year, to replicate past research which has nominated the number of student failures or low grades as a predictor of student dropout (Alexander et al., 2001; Allensworth & Easton, 2005; Balfanz et al., 2007). Acknowledging the previous research that has shown that one of the most powerful predictors of student dropout is retention (Jimerson et al., 2002; Jimerson et al., 2005; Roderick, 1994;

Rumberger, 1995), the variable Retained includes all students who were ever retained at any grade level. As discussed above, non-cumulative GPA is used rather than overall GPA as the average of the student's grades across all subjects within each grade level.

A discrete-time hazard model was fitted to the data by estimating parameters for each time period and for each of these variables using logistic regression. However, because no students dropped out before grade 7; intercept parameters for grades 1 through 6 are not determinable, and thus not included in the model. Hence, the beginning of time for the model is grade 7. In contrast to OLS regression techniques, to include the conditional effect of time in estimating the risk of experiencing an event of interest at any one time-point<sup>3</sup>, discrete-time hazard models usually begin with a test for the significance of multiple pseudo "intercepts" for each time-point, modeling the effect of time in the analysis of a subject's risk of the event. Additional parameters that estimate the effects of variables collected on a sample are then fit to the model as  $\beta$  estimates, similar to OLS regression, and then model fit is assessed. Seven discrete-time hazard models are presented in Table 4, listing parameter estimates and significance levels, standard errors for each estimate (in parentheses), the overall *N* of the person-period dataset at each time-point, and tests of model goodness of fit, including -2Log-likelihood, Chi-Square, and Cox & Snell pseudo R<sup>2</sup>.



The first model fitted to the data included only time-point intercepts for Grades 7 through 12 (*Table 4, Model A*). Each time-point parameter was significant, and the model appears to moderately fit the data according to the goodness of fit statistics. This indicates that for the base model including only time as a predictor, as would be expected from the trends plotted in Figure 1, each grade level, beginning at grade 7, is significant when considering a student's risk of dropping out. Model B builds upon Model A, in an attempt to hold known categorical predictors of dropout constant in the equations, and included significant parameter estimates for the time invariant variables Gender, Ethnicity and District (*Table 4, Model B*). Because both retention and the number of low and failing grades are known to be associated with students not graduating on time, Model C included the main effects of Retained as a dichotomous time varying variable (repeated a grade level = 1, never retained = 0) and DETotal (total number of D's or less per grade level). In Model C, all of the parameter estimates are significant, and the -2Log-likelihood decreased from Models A and B, indicating less remaining variance left unexplained, while the Chi-Square increased, indicating that Model C fits the data better than Models A and B.

Having thus tested and replicated previously identified predictors of student dropout, this discussion now turns to the main question of this study. The question of interest concerns the main effect of teacher assigned grades on a student's risk of not graduating on time. To test grades in the model, the variable "Non-cumulative GPA" (within-year non-cumulative grade point average for all subjects taken that year) was added (*Table 4, Model D*). The addition of Non-cumulative GPA radically shifted the estimates and significance levels of the majority of the parameters in the model, reducing or eliminating the significance of each time-point, as well as the time invariant variables Gender, Ethnicity and District. As the main significant finding of this study, in support of the idea that teacher assigned grades are a significant predictor of student dropout, this finding suggests that the variable Non-cumulative GPA accounts for much of the variance in the estimated probability of a student dropping out of school, more so than grade level, gender, ethnicity, district, or DETotal.

To explore the fit of more parsimonious models, Models E, F and G were estimated. Model E fits to the data only those variables that had significant parameters in Model D, or represented a continuous stretch of time from grades 9 through 12. Model F fits only the most significant parameters, Grade 11, Retained and Non-cumulative GPA. Model G, as a subset of Model F, fits only Retained and Non-cumulative GPA. While the difference between models F and G is statistically significant (change in -2Log-Likelihood = 9.64 ( $df=1$ )  $p < 0.01$ ) the magnitude of the effect of the Grade 11 time point on the model is weak, as evidenced by the relatively small 0.2% difference in the pseudo  $R^2$  results between the two models, and thus contribution to the variance explained (*Table 4*). Hence, to simplify analysis and interpretation while using the most parsimonious model, the remaining results and discussion will focus on Model G.

As the final model, Model G is interesting in four ways. First, it is acknowledged that all logistic regression pseudo- $R^2$  calculations are notoriously inaccurate as they approach 1.0, due the issue that no true  $R^2$  calculations exist for logistic regression analysis (Aldrich & Nelson, 1984). Hence, for all of the models presented here, once the  $R^2$  calculation, as an indicator of the amount of variance explained by the equation, surpasses 0.5 one must be

cautious in the interpretation of the accuracy of that calculation. However, Model G appears to explain well over 50% of the variance in the probability of a student dropping out of school (Cox & Snell  $R^2 = 0.707$ ), which is an improvement over past logistic regression dropout estimations (Alexander et al., 2001; Balfanz et al., 2007). In addition, Model G is reasonable given past research that nominated retention and grades as useful in identifying dropouts. Second, Model G does not contain any time-point parameters, suggesting that at any grade level a student's risk of dropout in this sample is explained well by that student's retention status and non-cumulative GPA. Stated another way, for this dataset, a student's risk of dropping out is mainly time-invariant, and is based more on a student's non-cumulative GPA and retention status. This is a significant finding when considering that past at-risk prediction and prevention measures have considered time to be significant, and grades to be only one of many possible variables to assess risk (Alexander et al., 2001; Balfanz et al., 2007; Bradley & Lenton, 2007; Eckstein & Wolpin, 1999; Jimerson et al., 2000).

Third, it is interesting that none of the three time invariant parameters were significant in the final model. Time invariant variables are usually included in discrete-time hazard models to help control for factors that lead to sample bias, in this study postulated to be gender, ethnicity and which district a student had attended. The inclusion of Non-cumulative GPA shifted the parameters for these variables to non-significant levels in Model D, suggesting that Non-cumulative GPA explains more of the variance in a student's risk of dropping out than a student's gender, ethnicity or which of the two districts the student attended. Fourth, transforming the logit parameter estimates in Model G into odds denotes that for this sample, when controlling for a student's non-cumulative GPA, at any grade level after grade 6, students who are retained at any time in a school district are 10.2 times more likely to dropout than students who are not retained<sup>4</sup>. Transforming the same Retained logit parameter estimate into hazard probabilities<sup>5</sup> indicates that retained students are 91.1% more likely to dropout than non-retained students, replicating past research and extending it to more precise estimates for the risk of dropping out using survival analysis. When controlling for retention, transforming the logit parameter estimate for Non-cumulative GPA to odds signifies that for every one unit increase in non-cumulative GPA, students are 0.161 times less likely to dropout. More intuitive is to invert this calculation, which indicates that at any grade-level after grade 6, for every one unit increase in non-cumulative GPA, one whole letter grade, students are 6.02 times more likely to graduate. Odds ratios such as these are difficult to interpret (Aldrich & Nelson, 1984; Davies, Crombie, & Tavakoli, 1998), thus more intuitive is to transform the logit parameter for Non-cumulative GPA into a hazard probability. This indicates that when controlling for if a student has ever been retained, in any one grade-level after grade 6 a one unit increase in non-cumulative GPA, one whole letter grade, corresponds to a 13.9% decrease in a student's risk of dropping out. Together, these results confirm the negative impact of retention found in previous studies, but more importantly, provide new evidence that suggests that teacher assigned grades, as recorded as non-cumulative GPA, are a significant and important predictor of a student's longitudinal risk of dropping out of school. In addition, it appears that this risk, rather than being restricted to the high school levels, begins in middle school.

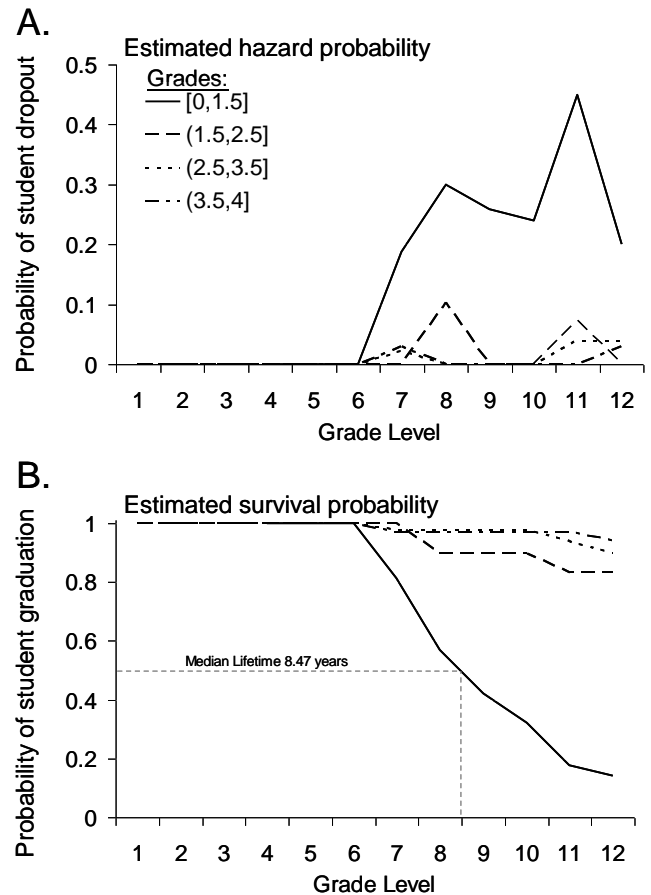
The final model, Model G, was tested for assumption violations of linear additivity and proportional hazard. The first assumption is of linear additivity, that the effect of predictors is linear. For this

study, one could imagine a difference in the behavior of low GPA versus high GPA students that may not necessarily be linear. To test this assumption, non-cumulative GPA was categorized at four levels,  $[0,1.5]$ ,  $(1.5, 2.5]$ ,  $(2.5,3.5]$ ,  $(3.5,4]$ , and parameters were tested in the model with retention. Overall, when retention is controlled for, non-cumulative GPA appears to generally behave in a linear manner in predicting student dropout, with increasing levels of non-cumulative GPA rising multiplicatively with a leveling off only between the top two categories (data not shown). The assumption of proportionality was assessed with the assumption that each predictor has an identical effect in every time period under study. This was assessed by evaluating interaction terms between each grade level, 7 through 12, with Non-cumulative GPA in the logit regression model predicting dropout. While it was found that each interaction term was significant and it also appeared that grades 9 and 10 may have a slightly larger interaction with non-cumulative GPA than the other grade levels, the overall model was less parsimonious and a much poorer fit than Model G (data not shown) and so proportionality over time is assumed. In addition, the deviance residuals were analyzed to explore how well the model performs for individual cases. Few cases had extremely high outlier deviance residuals (data not shown). Thus, overall, Model G appears to be well specified, fits the data well, and does not appear to violate any major assumptions of discrete-time hazard modeling.

In reference to a final issue of unobserved heterogeneity, which could pose a problem for a discrete-time hazard model such as this due to the possibility that one or more important predictors have been omitted from the equation that could explain the risk profile identified, the recommendations of Singer and Willet (2003) have been followed. Because unobserved heterogeneity asserts a consistent effect over time that leads to hazard functions that decline, since the hazard functions for this study all rise substantially over time, unobserved heterogeneity is not considered problematic. Additionally, the aim of this study is to identify if teacher assigned grades, as represented by non-cumulative GPA, are a significant and useful predictor of dropout that should be considered as useful by school districts in identifying students at-risk of dropping out. The aim is not to identify a single predictor equation that is generalizable to a larger population, such as attempting to infer that Model G could represent a population-level estimation equation of student dropout, the sample size is insufficient to make this claim. Thus, even if unobserved heterogeneity is an issue here, which it most likely is not due to the increasing risk profile over time, that would not negate the point that Model G suggests that non-cumulative GPA is a significant and important variable to consider when predicting student dropout; the main focus of this study.

#### *Interpreting the Results of the Discrete-Time Hazard Model*

Because the hazard model in Model G is mainly time invariant, predicting a student's risk of dropping out after grade 6 solely on retention status and non-cumulative GPA, it is not possible to plot the fitted model over time. However, the model is striking in the specification of the role that non-cumulative GPA plays in the risk of student dropout. Thus, to provide an intuitive graphical display for results interpretation, the estimated sample hazard and survival functions were disaggregated for non-cumulative GPA and plotted (Figure 2, A and B).



**Figure 2:** *Estimated hazard and survival functions, disaggregated by non-cumulative GPA categories.* Non-cumulative GPA was divided into four categories and hazard and survival functions were estimated and plotted. The subset of students most at risk of dropout was students who received non-cumulative GPAs from 0 to 1.5 (Panel A). Student risk began in grade 7. Risk at each grade level is highest for students with low grades, peaking in grade 11, with almost 45% of students with low grades at risk of dropout. The estimated survival function shows that students in the lowest non-cumulative GPA category had the lowest rates of survival, with a median lifetime of 8.47 years, and only 14% surviving in school to graduation in grade 12 (Panel B).

Plotting the estimated hazard and survival functions by non-cumulative GPA category shows that students who received the lowest grades were most at risk of dropping out. Non-cumulative GPA was divided into four categories corresponding to the four major letter grades of E or F through D, C, B and A,  $[0,1.5]$ ,  $(1.5,2.5]$ ,  $(2.5,3.5]$ ,  $(3.5,4]$  respectively. Hazard and survival functions for dropout were estimated and plotted for students in these four categories (Figure 2). The hazard function indicates that students in the lowest non-cumulative GPA category were at the highest risk of dropout in every grade-level after grade 6 (Figure 2, A). Periods of highest risk for this group were in grade 8 (30% of students with low grades dropped out), and grade 11 (45% of students with low grades dropped out). Additionally, the survival function shows that students who received the lowest grades had the lowest rates of on-time graduation (Figure 2, B). The median lifetime indicates that the average student who receives grades in the lowest category stays in school for 8.47 academic years,

suggesting that the average student in the lowest non-cumulative GPA category drops out of school before the start of the second semester of grade 8. Additionally, the survival analysis shows that only 14% of the students in the lowest non-cumulative GPA category graduate on time by grade 12 (Figure 2, B). Moreover, these graphical results compare well with the results of Model G and the overall hazard estimates (Table 4). Together, they provide further evidence that non-cumulative GPA explains much of the variance in the probability of a student dropping out of school since the overall hazard trend for the dataset is highly similar to the hazard trend for student dropout from the low non-cumulative GPA category (compare Figures 1 and 2). These results are novel and significant, calculating risk of dropout based on teacher-assigned grades, utilizing the entire grading scale, all subjects for each grade-level, a novel calculation of grades with non-cumulative GPA, all grade levels rather than just middle school or just high school, and handling the data appropriately using a life table and discrete-time hazard modeling. It appears that teacher assigned grades are highly predictive of a student's risk of dropping out, and suggests that grades are useful and should play a much larger role in the prediction of student "at-risk" status. Additionally, in comparison with previous methods of predicting student risk of dropping out, Gleason and Dynarski (2002) showed that a regression composite of multiple risk factor variables only accurately predicted 42% of the students who would have dropped out. More recently, Balfanz et al. (2007) were able to identify up to 60% of the students who drop out, using information from the end of grade 6, including a failing grade in English or mathematics. The results presented here in Figure 2B indicate that 86% of the students who received low grades, as measured by non-cumulative GPA, did not graduate on time. This appears to be a significant improvement over past at-risk prediction methods.

## DISCUSSION

In the longitudinal grade 1 through 12 analysis for the event of dropping out of school, this study has come to six main conclusions. 1) The use of survival analysis and life tables in studying dropout appears to be useful and informative. At the minimum, this study affirms that life tables and survival analysis that utilize student-level data from grades 1 through 12 may be an improvement over past graduation and dropout calculation methods. 2) It appears that risk of dropout in this dataset begins in grade 7 in middle school, two years earlier than the majority of current district at-risk prediction and prevention programs begin. 3) The most hazardous years for dropout in this dataset appear to be grades 8 and 11, the transition before entering high school in grade 9, and the year when students are old enough to drop out of school legally. 4) Replicating past research, retention of students at any grade level is shown in this study to have a highly negative impact on a student's probability of graduating on-time. 5) Teacher assigned grades, as measured by non-cumulative GPA, appear to be a significant and useful predictor of student dropout, with students who receive grades in the lowest category also experiencing a drastically increased risk of dropping out. 6) And finally, in comparison to past research on grades and dropout, the methods detailed here appropriately control for both the longitudinal and conditional discrete-time nature of grades 1 through 12 student data when examining the utility of teacher assigned grades for predicting student dropout.

These results appear to be novel and significant; however, issues of validity and generalizability must be addressed. The sample size for this study is small, consisting of only two cohorts of students, from two school districts. This may have led to sample bias,

Bowers (2010) *Grades and Graduation*

district effects, or cohort effects due to the intact nature of the sample. However, this issue is attenuated somewhat by the inclusion of two school districts in the study as well as the power analysis which indicated that the sample size is sufficient. Additionally, the longitudinal nature of the study increases its internal validity for these two districts. Overall, while this issue of an intact sample must be addressed, the findings of this study correspond to the findings of similar studies in different locations, and extend those findings to a quantification of low grades as a significant and useful predictor of student dropout, through an initial test of the use of discrete-time hazard modeling using non-cumulative GPA. Future studies should include many more cohorts of students in many more districts to help control for this issue as well as further explore the utility of survival analysis in predicting the risk of student dropout.

The main question of this study was to investigate teacher assigned grades as a predictive indicator for students at risk of dropout through an initial use of life tables, survival analysis and discrete-time hazard modeling. The results show that for this dataset, grades as measured by non-cumulative GPA, are predictive of students at risk of dropping out, and that this risk is greatest for students who receive the lowest grades. Grades were a major contributor to the fit of the discrete-time hazard model, outperforming previously known predictive categorical variables, such as gender, ethnicity, and even district effects. The predictive power of grades is made more evident when one compares the similarity in the shapes of the estimated hazard functions between Figures 1A and 2A. For this dataset, grades appear to account for much of the variance in risk of student dropout. Additionally, the use of grades to predict if a student will become at-risk appears to be an improvement over past methods, indicating that 86% of the students who received grades in the lowest category did not graduate on time. This suggests that for districts and schools wishing to assess if a student is at-risk of dropping out, that student's longitudinal grade history should be considered as a predictor of risk.

Nevertheless, the point that grades are important in predicting graduation may seem intuitive, if not trivial or banal. Shouldn't higher grades predict graduation? The argument here is that grades do predict graduation, but this point is important because according to the literature, schools currently do not use grades as data for decision making in the manner suggested in this study. Grades are seen as "hodge-podge" (Cizek, 2000; Cross & Frary, 1999), incorporating an assessment of not only academic knowledge, but also attendance, behavior, and participation. Also, grades are perceived in the literature as needing to be reported only to parents and students (Shepard et al., 2005). However, in the studies that have examined grades and student dropout, while methodological issues have somewhat clouded the point (as discussed above), with some studies concentrating on certain subjects (Eckstein & Wolpin, 1999), grade levels (Balfanz et al., 2007), or overall counts of failing grades (Allensworth & Easton, 2005), the overall conclusion that is also supported strongly by this study, is that teacher assigned grades are a significant predictor of student dropout. This is especially important given that schools already assign grades to students, rather than needing to add yet another new test to the school routine, and that grades have a high level of face validity with students, parents and teachers (Farr, 2000; Hargis, 1990; Kirschenbaum, Napier, & Simon, 1971; Shepard, 2006).

If grades are hodge-podge, and thus not a pure indicator of a student's academic knowledge, then how is it that low grades

appear to predict a student's risk of dropout? If low grades are predictive of students dropping out, one argument would be to give all students higher grades to prevent them from dropping out. However this would be confusing correlation with causation, and most likely would not lead to any reduction in student risk based on grades. An alternative argument however, is that while grades may not be a pure assessment of student's academic knowledge, because grades incorporate all of the factors indicated by the hodge-podge grading literature, grades may be an accurate assessment of a student's ability to negotiate the intricacies of the schooling process (Bowers, 2009). Teachers may be assessing a student's ability at this process through the grades they assign. Since teachers incorporate into grades whether or not a student attends class, participates, and hands in homework, as indicated by the hodge-podge grading literature, this study hypothesizes that teachers, through grades, may be accurately assessing a student's ability to perform well in the school process, as indicated by non-cumulative GPA, and thus graduate on time. Future studies will look more closely at this issue.

The discrete-time hazard model indicated that retention and non-cumulative GPA are more significant in predicting a student's risk than time invariant categorical variables, such as gender, ethnicity and district attended. As has been postulated in the past (Catterall, 1998), this shows that the use of a student's performance to date in the system provided to that student to predict at-risk status, rather than strictly relying upon categorical factors outside of the school context, such as family socio-economic status, is important and worthwhile. Much of the current practice in schools that attempts to determine a student's risk of dropout (Gleason & Dynarski, 2002; Hammond et al., 2007) may be overly biased towards such categorical variables, while ignoring the rich set of data that exists for each student at every grade level to date, assessing that student's performance in the system that's been provided to the student. To this end, future work will analyze the effects on risk of dropout of multiple assessments throughout a student's career in a school district, including grades and standardized assessments.

As suggested previously, but to date rarely utilized, (Bradley & Lenton, 2007; Willett & Singer, 1991) survival analysis and discrete-time hazard modeling appear to be novel and interesting methods for assessing the magnitude of the probability of the risk of student dropout at every grade level in which students leave school, in this study beginning after grade 6. However, for studies of this type, there is the question of using a single risk model over a competing risks model. In a single risk model, only one dichotomous event outcome is assumed (Singer & Willet, 2003), here student dropout versus graduation. It could be argued that a competing risks model could be used for this dataset, modeling three events, a student's probability of dropout, graduation or transfer into or out of a school district. However, the focus of this study is to understand the effect of grades on the prediction of student dropout, so extending the analysis to a competing risks model, while interesting, would over complicate the analysis as well as address a different research focus, modeling student transfer as an alternative to dropping out, and thus must be left to future studies. Moreover, for this study it is argued that a competing risks model is inappropriate in that the event under consideration is the termination of a student's school career. Thus, as a student transferring to another school cannot be considered a competing risk for such an event, since the student is still enrolled in a school somewhere, a single risk model is more appropriate with student transfer handled through censoring, as was done here.

For student dropout overall, the results of the survival analysis indicate that the most hazardous grade levels are grades 8 and 11. This replicates past research that has shown that the transition from middle school to high school is an especially hazardous time (Catterall, 1998; Eckstein & Wolpin, 1999; Roderick, 1993; Roderick & Camburn, 1999; Rumberger, 1995; Zvoch, 2006). It also shows that for the state in which these districts reside, student risk of school dropout increases in grade 11 after they have reached the age to legally dropout of school for the state, age 16. This point that a large percentage of students dropout during the middle school grades is significant for three reasons. First, because these results show that the median grade-level for dropout for students with low grades is at the grade 8 level, national and state-level graduation rate estimates that do not include the middle school grades (Swanson, 2004a) are missing an important segment of each district's student population that may have dropped out of school before grade 9. Second, these results further stress the need for longitudinal data on each student that spans at least the middle and high school grades, rather than concentrating only on the high school level or relying entirely on cross-sectional data, to better track and understand when and how students are at the greatest risk for dropping out of school. Third, as discussed above, the literature indicates that most at-risk prevention and intervention measures take place at the high school level (Dynarski & Gleason, 2002). The literature further indicates that a student's decision to dropout is not based on a single event, but rather builds from a long history of events (Alexander et al., 2001; Jimerson et al., 2000), eventually convincing the student to leave school early. The results presented here show that students begin to experience a risk of leaving school at the middle school level, which generally rises over subsequent years. Together, these issues point to the conclusion that at-risk identification and prevention measures must begin to be utilized much earlier than high school, starting at least at the middle school level.

Thus, this discussion leads to the question that if grades are predictive of student risk of dropout, starting in middle school, what is to be done about it? The results of this study do not speak to this issue. While outside the scope of this study, it is important to take up this question since accurate identification is only the first step of many in helping to address the needs of students who may be experiencing difficulties with school. However, to date, little work has been done to systematically evaluate at-risk prevention programs.

For most of the evidence on at-risk and dropout prevention, methodological problems persist which inhibit a robust evaluation of what works, such as biased groupings and estimates of effects, since randomized controlled trials are rarely performed in this area (Agodini & Dynarski, 2004; Lehr, Hansen, Sinclair, & Christenson, 2003; Rumberger, 2004a). Nevertheless, what the literature indicates is that historically, most dropout prevention programs appear to not reduce student dropouts (Dynarski, 2004; Dynarski & Gleason, 2002). As reviewed by Dynarski and Gleason (2002) and Lehr *et al* (2003), these programs mostly occur at the high school level and consist of helping students build self-esteem, overcome personal and family issues and increase attendance through periodic counseling; consist of the creation of smaller school settings; or provide tutoring or mentoring services. Similar programs at the middle school level have had somewhat more of an impact, but as discussed above, the accuracy of identification of students at risk of dropping out using middle school level data to date has been low and problematic. Hence, any program that appears to work using middle school level data, may have

“worked” only to the extent that the majority of the students identified for at-risk interventions were mis-identified originally as being students at risk of dropping out.

Acknowledging that much more high-quality work is needed in the evaluation of dropout prevention programs before any one individual program can be recommended over another (Dynarski & Gleason, 2002; Hammond et al., 2007; Lehr et al., 2003), recent literature has begun to urge for a shift from a deficit model of attempting to prevent dropouts, to a more positive model of promoting and encouraging successful school completion (Christenson & Thurlow, 2004). While definitive data on such programs is currently lacking, it will be interesting to combine increasingly accurate dropout prediction methods, such as the results presented here, with controlled studies evaluating the effects of providing resources to students to help them complete high school on time.

In the end, as discussed above, this study replicates and extends previous findings that have nominated teacher assigned grades as useful in predicting student dropout. For the first time, this study has shown with life tables, survival analysis and discrete-time hazard modeling, that a novel calculation of grades, non-cumulative GPA, examined from grades 7 through 12, is a significant and useful predictor of student dropout. Based on these findings, the author recommends that school districts begin to immediately investigate the utility of non-cumulative GPA as a very easy and cost effective number to calculate, in comparison to additional tests or surveys, as a primary means to identify students at risk of dropping out of school, starting at grade 7.

### RECOMMENDED CITATION:

Bowers, A.J. (2010) Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *The Journal of Educational Research*, 103(3), 191-207.  
[doi:10.1080/00220670903382970](https://doi.org/10.1080/00220670903382970)

### ENDNOTES

1. Multiple school districts and student cohorts were considered for this study, however some districts were not willing to participate and some districts did not retain data on students who dropped out of school, thus limiting the overall size of the study.
2. As recommended for data of this type, censoring indicates that data for a specific student includes all of the variables up until the time of the student's exit from the dataset, but with 0 recorded for the variable “dropout” rather than 1. For a review of censoring, see Singer & Willet (2003).
3. Each time-point is conditional due to the fact that any student who experiences the event in one time-point is removed from the calculations for all future time-points.
4. Odds =  $e^{\beta} = e^{(2.325)} = 10.2$
5. Probability =  $1/(1+e^{-\beta})$
6. A bracket indicates “inclusive” in a series of numbers, an open parenthesis indicates “non-inclusive”.

### ACKNOWLEDGEMENTS

The author would like to thank Susan Printy for her suggestions and critiques of this study.

### REFERENCES

- Abrams, L., & Haney, W. (2004). Accountability and the grade 9 to 10 transition: The impact on attrition and retention rates. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 181-205). Cambridge, MA: Harvard Education Press.
- Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *The Review of Economics and Statistics*, 86(1), 180-194.
- Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models*. Newbury Park, CA: Sage Publications.
- Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, 103(5), 760-822.
- Allensworth, E. M. (2005). Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004. <http://www.consortium-chicago.org/publications/p75.html>
- Allensworth, E. M., & Easton, J. Q. (2005). The on-track indicator as a predictor of High School graduation. <http://www.consortium-chicago.org/publications/p78.html>
- Balfanz, R., Herzog, L., & MacIver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.
- Balfanz, R., & Legters, N. (2004). Locating the dropout crisis: Which high schools produce the nation's dropouts? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 57-84). Cambridge, MA: Harvard Education Press.
- Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research*, 82(6), 309-319.
- Berkold, J., & Carroll, C. D. (1998). *Subsequent educational attainment of high school dropouts*. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=98085>.
- Bowers, A.J. (2007) Grades and data driven decision making: Issues of variance and student patterns, unpublished PhD thesis, Michigan State University, East Lansing, MI.
- Bowers, A.J. (2009) Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, 47(5), 609-629.
- Bradley, S., & Lenton, P. (2007). Dropping out of post-compulsory education in the UK: An analysis of determinants and outcomes. *Journal of Population Economics*, 20(2), 301-328.
- Cameron, S. V., & Heckman, J. J. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics*, 11(1), 1-47.
- Catterall, J. S. (1998). Risk and resilience in student transitions in high school. *American Journal of Education*, 106(2), 302-333.
- Christenson, S. L., & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science*, 13(1), 36-39.
- Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, 19(2), 16-33.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.

- Davies, H. T. O., Crombie, I. K., & Tavakoli, M. (1998). Information in practice: When can odds ratios mislead? *British Medical Journal*, *316*, 989-991.
- Dynarski, M. (2004). Interpreting the evidence from recent federal evaluations of dropout-prevention programs: The state of scientific research. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 255-268). Cambridge, MA: Harvard Education Press.
- Dynarski, M., & Gleason, P. (2002). How can we help? What we have learned from recent federal dropout prevention evaluations. *Journal of Education for Students Placed at Risk*, *2002*(1), 43-69.
- Eckstein, Z., & Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, *67*(6), 1295-1339.
- Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education*, *65*(2), 91-113.
- Farr, B. P. (2000). Grading practices: An overview of the issues. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 1-22). Norwood: Christopher-Gordon Publishers.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117-142.
- Fitzsimmons, S. J., Cheever, J., Leonard, E., & Macunovich, D. (1969). School failures: Now and tomorrow. *Developmental Psychology*, *1*(134-146).
- 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*(1), 25-41.
- Greene, J. P., & Caire, K. (2001). High school graduation rates in the United States. [http://www.manhattan-institute.org/pdf/cr\\_baeo.pdf](http://www.manhattan-institute.org/pdf/cr_baeo.pdf)
- Greene, J. P., & Winters, M. A. (2005). Public high school graduation and college-readiness rates: 1991-2002. [http://www.manhattan-institute.org/html/ewp\\_081.htm](http://www.manhattan-institute.org/html/ewp_081.htm)
- Hammond, C., Linton, D., Smink, J., & Drew, S. (2007). *Dropout risk factors and exemplary programs: A technical report*. Clemson, S.C.: National Dropout Prevention Center.
- Hargis, C. H. (1990). *Grades and grading practices: Obstacles to improving education and helping at-risk students*. Springfield: Charles C. Thomas.
- Heckman, J. J., & LaFontaine, P. A. (2007). *The American high school graduation rate: Trends and levels*. Bonn, Germany: Institute for the Study of Labor.
- Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the schools*, *39*(4), 441-457.
- Jimerson, S. R., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, *38*(6), 525-549.
- Jimerson, S. R., Pletcher, S. M. W., Graydon, K., Schnurr, B. L., Nickerson, A. B., & Kundert, D. K. (2005). Beyond grade retention and social promotion: Promoting the social and academic competence of students. *Psychology in the schools*, *43*(1), 85-97.
- Kienzi, G., & Kena, G. (2006). *Economic outcomes of high school completers and noncompleters 8 years later* (No. NCES 2007-019). Washington D.C.: National Center for Education Statistics.
- Kirschenbaum, H., Napier, R., & Simon, S. B. (1971). *Wad-ja-get? The grading game in American education*. New York City: Hart Publishing Company.
- Laird, J., DeBell, M., & Chapman, C. (2006). *Dropout rates in the United States: 2004*. Washington, D.C.: National Center for Education Statistics.
- Lakatos, E. (1988). Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*, *44*(1), 229-241.
- Lehr, C. A., Hansen, A., Sinclair, M. F., & Christenson, S. L. (2003). Moving beyond dropout towards school completion: An integrative review of data-based interventions. *School Psychology Review*, *32*(3), 342-364.
- Lloyd, D. N. (1974). Analysis of sixth grade characteristics predicting high school dropout or graduation. *JSAS Catalog of Selected Documents in Psychology*, *4*, 90.
- Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, *38*(4), 1193-1200.
- Marrow, G. (1986). Standardizing practice in analysis of school dropouts. *Teachers College Record*, *87*(3), 342-355.
- Montes, G., & Lehmann, C. (2004). *Who will drop out from school? Key predictors from the literature* (No. T04-001). Rochester, NY: Children's Institute Inc.
- NCES. (2004). *National Institute of Statistical Sciences/Education Statistics Services Institute task force on graduation, completion, and dropout Indicators*. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005105>.
- NCES. (2006). Common Core of Data. <http://nces.ed.gov/ccd/>
- Orfield, G. (2004). Losing our future: Minority youth left out. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 1-11). Cambridge, MA: Harvard Education Press.
- Randolph, K. A., & Orthner, D. K. (2006). A strategy for assessing the impact of time-varying family risk factors on high school dropout. *Journal of Family Issues*, *27*(7), 933-950.
- Roderick, M. (1993). *The path to dropping out: Evidence for intervention*. Westport, CT: Auburn House.
- Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, *31*(4), 729-759.
- Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of High School. *American Educational Research Journal*, *36*(2), 303-343.
- Roderick, M., Nagaoka, J., Bacon, J., & Easton, J. Q. (2000). Update: Ending social promotion. <http://www.consortium-chicago.org/publications/pdfs/p0g01.pdf>
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, *32*(3), 583-625.
- Rumberger, R. W. (2001). *Why students drop out of school and what can be done*. Paper presented at the conference: Dropouts in America: How Severe is the Problem? What Do We Know about Intervention and Prevention?, Harvard University.
- Rumberger, R. W. (2004a). What can be done to reduce the dropout rate? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 243-267). Cambridge, MA: Harvard Education Press.
- Rumberger, R. W. (2004b). Why students drop out of school. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 131-156). Cambridge, MA: Harvard Education Press.
- S&P. (2006). School Matters. <http://www.schoolmatters.com/>

- Sable, J., Gaviola, N., & Hoffman, L. (2007). *Numbers and rates of public high school dropouts: School year 2004-05. First Look* (No. 2008-305). Washington D.C.: National Center for Education Statistics.
- Seastrom, M., Hoffman, L., Chapman, C., & Stillwell, R. (2006). *The averaged freshman graduation rate for public high schools from the common core of data: School years 2002-03 and 2003-04*. <http://nces.ed.gov/pub2006/200606rev.pdf>.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (Fourth ed.). Westport, CT: Praeger Publishers.
- Shepard, L. A., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J. B., Gordon, E., et al. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass.
- Singer, J. D., & Willet, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), 155-195.
- Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Swanson, C. B. (2004a). Sketching a portrait of public high school graduation: Who graduates? Who doesn't? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 13-40). Cambridge, MA: Harvard Education Press.
- Swanson, C. B. (2004b, Feb). Who graduates? Who doesn't? A statistical portrait of public high school graduation, class of 2001. [http://www.urban.org/UploadedPDF/410934\\_WhoGraduates.pdf](http://www.urban.org/UploadedPDF/410934_WhoGraduates.pdf)
- Tyler, J. H. (2003). Economic benefits of the GED: Lessons from recent research. *Review of Educational Research*, 73(3), 369-403.
- Viadero, D. (2006). Signs of early exit for dropouts abound. *Education Week*, 25(41S), 20-22.
- Weber, J. M. (1989). *Identifying potential dropouts: A compilation and evaluation of selected procedures*. Columbus, OH: Ohio State University.
- Willet, J. B., & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research*, 61(4), 407-450.
- Willet, J. B., & Singer, J. D. (1995). It's deja vu all over again: Using multiple-spell discrete-time survival analysis. *Journal of Educational and Behavioral Statistics*, 20(1), 41-67.
- Zvoch, K. (2006). Freshman year dropouts: Interactions between student and school characteristics and student dropout status. *Journal of Education for Students Placed at Risk*, 11(1), 97-117.