# At What Point Do Schools Fail to Meet Adequate Yearly Progress and What Factors are Most Closely Associated with Their Failure? A Survival Analysis[1,2]

Penny L. Pruitt
The University of Texas at San Antonio

Alex J. Bowers[3]
Teachers College, Columbia University

## ABSTRACT:

The purpose of this study is to examine the factors most associated with the probability of Texas high schools failing to make Adequate Yearly Progress (AYP) under NCLB, examining the entire population of all Texas public and charter high schools from 2003-2011, *n*=1721. While literature to date focuses on the different variables that may affect schools in meeting AYP as well as addressing the success and failure of charter schools, there is a lack of research on the specific variables that have the most impact on failing to meet AYP, considered here as a "hazard". We used discrete time hazard modeling to estimate the probability of a school failing AYP for the first time in the time period. Our findings indicate that rural schools were the least likely to fail AYP, while schools with higher percentages of African American and Hispanic students and larger class sizes and enrollments failed AYP much more often. As a components of the AYP state determinations, school percent met standard in mathematics and attendance were significant in the model, while graduation rates were not. This study provides one of the first opportunities to examine the year-by-year hazard of failing AYP in Texas over the first decade of NCLB implementation.

**Key Words:** attendance rate, Adequate Yearly Progress, adequacy, assessment, economically disadvantaged, graduation rate, hazard function, No Child Left Behind Act (NCLB) survival probability.

[1] This document is a manuscript presented at the 2014 annual meeting of the Association of Education Finance and Policy (AEFP).
[2] Citation: Pruitt, P.L.,Bowers, A.J. (2014) At What Point Do Schools Fail to Meet Adequate Yearly Progress and What Factors are Most Closely Associated with Their Failure? A Survival Model Analysis. A paper presented at the annual meeting of the Association of Education Finance and Policy, San Antonio, TX: March 2014.
[3] Alex J. Bowers (bowers@tc.edu); Teachers College, Columbia University; Bowers@tc.edu; 525 W. 120th Street, New York, New York 10027. ORCID: 0000-0002-5140-6428, ResearcherID: C-1557-201

## INTRODUCTION

The purpose of this study was to analyze the factors most associated with the probability of Texas high schools failing to meet Adequate Yearly Progress (AYP) during the first decade of No Child Left Behind (NCLB). This study focuses on the hazard probability of failing to meet AYP and the factors that most influence the failure to meet AYP utilizing data from the Texas Education Agency (TEA) for the years 2003 – 2011 and the National Center for Education Statistics (NCES) Common Core of Data.

Although both secondary and elementary campuses are required to meet AYP in Texas during this time period, this study only focused on high school campuses. While NCLB was designed to identify schools that fail to meet the educational needs of low-income and minority students, it was also designed to identify high schools that consistently have high numbers of students who do not perform at a proficient academic level and do not graduate from high school within four years. While literature to date identifies numerous variables affecting high school failure to meet AYP, the challenges schools and districts face in meeting the requirements of NCLB, and the impact of failing to meet AYP, there is no current research on the specific variables that have the most impact on failing to meet AYP (Balfanz, Legters, West, & Weber, 2007; Lee & Reeves, 2012; Linn, 2003; Mintrop & Sunderman, 2009; "NCLB," 2002).

## BACKGROUND AND LITERATURE REVIEW

Pressure from business and political leaders to develop an accountability system gained strength in the 1980s after the publication of *A Nation at Risk* underscored how poorly American high school students performed on international assessments (U.S. Department of Education & The National Commission on Excellence in Education, 1983). What *A Nation at Risk* failed to emphasize was that results were unreliable because comparisons were not based upon equal cohorts (Berliner & Biddle, 1995). Furthermore, the authors failed to highlight that the difference in average scores was quite small. Whereas an average score of 38% earned the United States a 10th place ranking, a 44% earned Scotland a 4th place ranking. Moreover, tests in algebra in 8th grade ranked American students dead last, yet most American students do not take algebra in 8th grade. When comparing the elite students who do take algebra in 8th grade to elite Japanese students in 8th grade, American students scored higher. Despite statistical evidence to the contrary, corporate America and government officials had a scapegoat for the economic woes of the country and began blaming a poor educational system for the economic problems of the country (Berliner & Biddle, 1995; Bracey, 1996; Hursh, 2007; Sloane & Kelly, 2003).

In response to pressure from business and political leaders, most states began developing curricular standards and utilizing standardized tests to assess achievement. Texas was one of the first states to develop statewide standardized tests in the 1980s, and it began requiring minimum competency testing for graduation in 1987. In 1993, the State developed the Public Education Information System (PEIMS) to track data and begin holding schools accountable. Schools were measured on the total number of students passing the Texas Assessment of Academic Skills (TAAS) reading and math assessments with a 70% or higher and the annual dropout or completion rate. Additionally, campuses had to meet the same requirements within the disaggregated by subgroup populations – White, African American, Hispanic, and economically disadvantaged. Campuses and districts were rated exemplary if 90% of the total students, as well as the disaggregate subpopulations, were proficient and the dropout rate did not exceed 1%. They were rated recognized if 5% of the total population and subgroups met the passing standard and less than a 3.5% dropout rate, and they were rated acceptable if 50% of the students and each subgroup were rated proficient and no more than a 6% dropout rate. Anything lower than these scores and dropout rates earned campuses and districts a low-performing rating. During the mid-late 1990s, the passing standard increased from 70% to 80%, and a new test was introduced. The Texas Assessment of Knowledge and Skills (TAKS) was no longer a minimum competency test, but one of increased rigor and high-stakes (Vasquez Heilig & Darling-Hammond, 2008).

Up until 2002 each state determined the significance of their assessments. In 2002, utilizing Texas as a model, President George W. Bush signed into law the reauthorization of the Elementary and Secondary Education Act of 1965 which became known as the *No Child Left Behind Act* (NCLB) (Hursh, 2005). With the implementation of NCLB in 2002, states and school districts have faced increasing pressure to meet federal accountability standards in order to continue receiving federal funding. If schools or districts receive federal funds, such as Title I funds due to the number of low-income students, they must meet Adequate Yearly Progress" (AYP) in order to maintain these federal funds. AYP measures include performance on state standardized tests in both reading and mathematics and another indicator. In Texas, attendance is the other indicator for elementary schools, and completion rate is the other indicator for high schools.

Meeting AYP means schools must have proficient scores by state standards in reading and math and in all subgroup populations, which includes White, African American, Hispanic, Asian, Native American, economically disadvantaged, limited English proficient, and special education students. If a school fails to meet AYP for two years in a row, regardless of the indicator in which they fail to meet the standard, students in that school must be given an option to transfer to another public school that has not been identified as needing improvement, and the district must provide students opportunities for supplemental services, such as tutoring provided by outside agencies, remedial classes, and summer school. In addition, if the school fails to meet AYP for a third consecutive year, the district must take corrective action. During the fourth year of AYP failure, the district must develop plans for restructuring the school, and if the school fails in the fifth year, the district is required to implement the restructuring, which may include being converted into a charter school, or being closed (Lee & Reeves, 2012; Linn, 2003; "NCLB," 2002).

Advocates of NCLB and standardized assessments argue that the implementation of such measures increases educational equality, while opponents, contend that the high stakes tests associated with NCLB have actually reduced educational standards because teachers are teaching to the tests or states have lowered the proficiency standards to allow students to meet the norm. Some research indicates that students in lower grade levels are performing better on the National Assessment of Educational Progress (NAEP) after the implementation of NCLB, thus reducing the achievement gap. Other studies, however, indicate that the achievement gap has, at best, fallen flat if not widened in high school (Amrein & Berliner, 2003; B. Fuller, Wright, Gesicki, & Kang, 2007; E. J. Fuller & Johnson, 2001; Haertel, 1999; Hursh, 2005; Lee & Reeves, 2012; McDill, Natriello, & Aaron, 1986; Wei, 2012).

While there is some evidence that accountability has benefited some minority and economically disadvantaged students, the evidence is thin and contradictory (Lauen & Gaddis, 2012). Two studies reported sizeable gains for fourth-grade students on math National Assessment of Educational Progress (NAEP) scores and smaller, but significant gains for eighth-grade math scores, while other studies indicate that there have been no gains and some regression in the achievement gap between Whites and minority students, particularly the Black-White achievement gap and predominantly in states with more rigorous achievement assessments. The largest test disparities occur in schools that are racially segregated, with Whites outperforming all groups and Hispanics outperforming

Blacks (Dee & Jacob, 2011; Hanushek & Raymond, 2005; Lee & Reeves, 2012; Price, 2010; Stiefel, Schwartz, & Chellman, 2007; Weckstein, 2003; Wei, 2012).

Many large school districts are attempting to meet AYP standards in high poverty schools by breaking them into smaller schools in an effort to improve educational inequalities (Orfield & Lee, 2005). Unfortunately, this leads to even more segregated schools (Orfield & Lee, 2005). Furthermore, attempting to create smaller schools exacerbates the already strained financial purse-strings many school districts face. Nonetheless, if the districts are unable to improve the performance of failing schools, they will be forced to provide supplemental educational services, including after-hours tutoring from inside or outside sources, and allow students to choose to attend schools that are not failing (Burch, Steinberg, & Donovan, 2007; DiMartino & Scott, 2013; Heinrich, Meyer, & Whitten, 2010). School choice for some districts could be devastating due to overcrowding or failing to meet their court-mandated desegregation orders. Insufficient capacity may not be used as a reason to not offer school choice; thus, districts are required to create additional capacity without funding to do so (2002). Although some districts have filed suit against the government to block the school choice mandate because NCLB choice will cause them to violate their court-ordered desegregation, a pre-standing court desegregation mandate does not exempt the district from NCLB compliance (DeBray-Pelot, 2007). According to NCLB, a district is required to secure a court's approval to change its desegregation plan. Failure to secure the court order puts a district out of compliance with Title I under NCLB. Yet, if they allow school choice, they are out of compliance with their court-ordered desegregation plan, leaving districts in a no-win situation (DeBray-Pelot, 2007).

NCLB also requires that states introduce "sanctions and rewards" relevant to every school and based on their AYP status. It mandates explicit and increasingly severe sanctions for persistently low-performing schools that receive Title I aid. Sanctions include anything from implementing public-school choice to staff replacement to school restructuring. According to data from the Schools and Staffing Survey of the National Center for Education Statistics, 54.4 percent of public schools participated in Title I services during the 2003–04 school year. Some states applied these explicit sanctions to schools not receiving Title I assistance as well. For example, 24 states introduced accountability systems that threatened all low-performing schools with reconstitution, regardless of whether they received Title I assistance (Dee & Jacob, 2010).

Some schools and districts are being met with the challenge of pulling resources from other areas to meet the mandates. Research indicates that school district expenditures have increased by nearly $600 per pupil due to direct student instruction and educational support services, yet the increased expenditure was not matched by federal funding support, nor was there increased test score gains associated with the increased expenditures (Dee, Jacob, M., & Ladd, 2010). The direct costs of managing an accountability system are quite small on a per-pupil basis. However, standards-based reforms have often been presented to the public as a trade: greater resources and flexibility for educators in exchange for greater accountability (Dee et al., 2010).

One of the most strident criticisms of NCLB is that it failed to deliver on this bargain (Dee et al., 2010). One noteworthy study on the relationship of spending and accountability is an analysis of district-level expenditure data from 1991-1997, produced by the National Center for Education Statistics (Hannaway, McKay, & Nakib, 2002). The study examined four states that implemented comprehensive accountability programs in the 1990s – Kentucky, Maryland, North Carolina, and Texas. They found that only Texas and Kentucky increased educational expenditure more than the national average, but those increases were substantial. An additional study suggests that the major pre-NCLB accountability reforms in Florida were associated with increased expenditure for instructional staff support and professional development, particularly in low-performing schools. What is not clear is whether the accountability policy caused the increased expenditure or whether both were merely parts of a broader reform agenda. Overall, the extant literature offers at best suggestive evidence on how accountability reforms may have influenced school spending (Dee & Jacob, 2010; Rouse, Hannaway, Goldhaber, & Figlio, 2007).

To provide new evidence on how NCLB influenced local school finances, Dee, Jacobs, and Schwartz (2010) pooled annual, district-level data on revenue and expenditure from U.S. Census surveys of school district finances over the period from 1994 to 2008. Their sample consisted of all operational, unified school districts nationwide for each survey year. To identify the effects of NCLB accountability on district finances, they utilized cross-state trend analysis, comparing within-state changes in school finance measures across states with and without pre-NCLB accountability programs. The results indicate that total per-pupil expenditure rose more quickly from 1994 to 2002 in states that adopted pre-NCLB accountability policies, and continued to grow, albeit more slowly, following the introduction of NCLB, suggesting that NCLB increased expenditure. The results indicate that NCLB increased total current expenditure by $570 per pupil, or by 6.8 percent from the 1999–2000. The increased expenditure was allocated to direct instruction and support services. Results also reveal that the increased expenditure was not matched by corresponding increases in federal support, consistent with allegations that NCLB constitutes an unfunded mandate. It should be noted, however, that the increase in spending on student support was not statistically significant. Furthermore, the effects were fairly similar across districts with different baseline levels of student poverty, suggesting that NCLB did not meaningfully influence distributional equity (Dee, Jacob, and Schwartz 2010).

Furthermore, under NCLB, states are required to withhold 2% of Title I funds for school improvement plans and local education agencies are required to hold and additional 10% of the federal funds they receive in reserve to provide for potential supplemental educational services instead of issuing it to schools that are most in need of the support. State education agencies also have the option of withholding an additional 5% of the funds in reserve for possible actions from the United States Department of Education (Johnson, 2007; United States Department of Education, 2002). Thus, schools and districts are being required to execute changes that cost them significantly with less money than they were originally allotted (Dee & Jacob, 2010; Johnson, 2007; Loveless, 2006).

In addition to schools and districts facing punitive financial measures enacted through NCLB, states are faced with the costs associated with implementing high stakes testing. Since 2010, New York awarded $33 million in contracts to Pearson Education to develop its states' standardized tests (DiMartino & Scott, 2013). Texas awarded approximately $90 million in contracts to Pearson Education to develop their new STAAR tests, and the costs will continue to rise with growth and the scoring of tests (Texas Education Agency, 2010). California's projected cost for the 2011-2014 standardized assessment contracts $167.5 million in creating and scoring the high-stakes tests that are required (California Department of Education, 2012). These funds could be applied to new programs aimed at truly improving education, yet they are going to private, for-profit industries, such as Pearson Education.

While states are spending millions of dollars in contracts to develop these standardized tests, many political leaders consider the use free-enterprise principles, such as school choice and private outside sources, will improve the educational system through accountability and development. The argument is that the public sector is too bureaucratic and disinterested in innovation. (DeBray-Pelot & McGuinn, 2009; DiMartino & Scott, 2013).

Proponents of charter schools claim they are innovative, yet their definition of innovation does not necessarily apply to classroom practices. There are very few studies that look at the classroom practices of charter schools, but the findings of those studies indicate that the classroom practices of charter schools are similar to those in public schools. Thus, the innovation in charter schools is based upon finance, organization, and autonomy. Additionally, it is not clear that charter schools are more effective in educating students than public schools. Some research indicates that test scores are improved when children attend charter schools, while other research suggests that charter school students are less successful than public school students (Bulkley & Fisler, 2003). In Washington D.C., more than 75% of students in 7 of the 16 charter schools scored below basic in math and reading. When compared to non-charter public schools in Washington D.C., test results were mixed but suggested that public charter schools were not achieving higher standards. In Texas, only 40% of the initial charter schools received ratings of academically acceptable or higher compared to 91% of public schools (Bulkley & Fisler, 2003; May, 2006). Yet, comparisons are difficult because the demographics of a charter school are not always similar to the surrounding public school. Furthermore, although charter schools are required to accept all students, they have been found to counsel parents of students with special needs to not enroll their child (Bulkley & Fisler, 2003; DiMartino & Scott, 2013; Henig, 1990; May, 2006).

Proponents of school choice also advocate the use of outside resources for remediation. NCLB mandates that schools who fail to make AYP for three consecutive years provide services to students to improve scores. While tutoring may be provided by school personnel after hours, supplement education services (SES) must also be offered through outside agencies at the schools' expense. Research indicates, however, that many of the private contractors used for SES are ineffective. Surveys show that many of the students who attend these outside supplemental educational services received very little instruction and virtually no one on one time aimed at improving their scores. Furthermore, instructional practices varied greatly between service providers, and even within the same provider, depending upon the setting and tutor. Although research suggests problems with these private contractors, the legislation strongly discourages any attempts by states and school districts to regulate tutorial services and the choices offered to parents (Heinrich et al., 2010).

With the implementation of NCLB, many state and local leaders were stunned at the number of their schools and districts that were previously recognized as successful schools being labeled as needing improvement or failing. Frequently, this is a result of not meeting standards within the subpopulations groups. Campuses that have few students have a far more difficult time of meeting the standard than campuses with larger populations (Fusarelli, 2004; Koyama, 2012; Novak & Fuller, 2003).

In order for districts and campuses to meet AYP on the reading/ELA and mathematics assessments, they must meet criteria based upon participation and performance. To meet the participation component, districts and campuses must have at least 95% of all students and student groups take the exams (Elledge, Le Floch, Taylor, & Anderson, 2009b).

Performance criteria requires districts and campuses to meet a specified passing rate each year for all student groups. Each year the specified pass rate increases, culminating in a 100% pass rate for all students by 2014. If campuses fail to meet the required performance measures, they may still meet AYP based upon required improvement or Safe Harbor. Safe Harbor is based upon two components. First, schools must show a 10% performance improvement for the student group on which they do not meet the standard and second they must also meet the criteria for the relevant other measure requirement for the student group. The performance improvement portion of the safe harbor calculation requires the calculation of actual change, defined as the current year number of students meeting the proficiency standard divided by the total number of students tested minus the previous year's number of students meeting the standard divided by the total number of students tested for that year. The actual change must be equal to or greater than the required improvement. If a campus was not measured on an item during the previous year and did meet the standard during the current year, they fail to meet AYP. In 2008, the United States Department of Education (USDE) approved an amendment to the requirement of the other measure in Safe Harbor for AYP that allows districts and campuses to meet the absolute standard for the other measure in order to satisfy required improvement. Thus, in the State of Texas, districts and campuses seeking to use required improvement must meet or exceed the graduation rate goal, annual targets, or alternatives for high schools and the attendance rate for elementary and middle schools must increase at least one-tenth of a percent (Texas Education Agency, 2012).

Moreover, districts and campuses missing the performance component on an indicator in one year but meeting it the next year may still fail to meet AYP if the participation component is not met. In such cases, the district or campus is considered to have missed AYP for that indicator two consecutive years, potentially triggering Title I school improvement program (SIP) requirements. The opposite also holds. If the district or campus misses participation on an indicator the first year, meets it the following year but misses performance for the same indicator, then the district or campus is again considered missing AYP for that indicator two consecutive years ("NCLB," 2002; Texas Education Agency, 2012).

Furthermore, NCLB legislation requires that states assess all students in reading/ELA including limited English proficient students. In Texas, English language learners may take the Texas English Language Proficiency Assessment System (TELPAS) test as part of the reading assessment. These results are used in lieu of the State of Texas Academic Assessment Reading (STAAR) system results for first-year or recent immigrants. Since the inception of STAAR, LEP students may take both TELPAS and STAAR-L (a linguistically accommodated version of STAAR), however, when this occurs, only the STAAR assessment results are used in the AYP reading calculation for that student. Decisions as to which test(s) the student should take are made by the language proficiency assessment committee (LPAC) (Texas Education Agency, 2012).

The other indicator used in Texas to measure AYP for high schools is graduation rate. The term graduates is based upon a longitudinal completion rate which is the percentage of students beginning in ninth grade who complete their high school education by the expected date. The longitudinal completion class has four components: 1) the percentage of students who complete high school by their expected graduation date or earlier; 2) the percentage of students who fail to complete by the end of four years but enroll in a public education institution the following year; 3) the percentage of students who earn a general educational development (GED) certificate; 4) the percentage of students who drop out of school (Texas Education Agency, 2012).

Not only do districts and campuses have to meet standards in assessment, participation, graduation, and attendance for the entire campus, they are required to meet the standards for all subgroups, including: English language learners, economically disadvantaged, special education, Hispanic, African American, Native American, White, and Asian. The student group is not included if the number of students in the subgroup is too small, as there is the potential to reveal personally identifiable information and the results would likely be statistically unreliable. Student group identifications are based on student characteristics and program participation used to report attendance rates for the state, and each state determines the number of students required to be included in the student group measure (B. Fuller et al., 2007; Fusarelli, 2004; Koyama, 2012; Linn, 2003; "NCLB," 2002; Texas Education Agency, 2012).

In Texas, for a student groups' measure to be evaluated for AYP, a student group must have 50 students and account for 10 percent of the student population or any group that has least 200 students will be evaluated, regardless of the percentage of population a district or campus. Furthermore, to be evaluated for AYP, a district or campus must have at least 9,000 days in membership for student groups of 50, 36,000 days for student groups of 200, or comprise at least 10 percent of total days in membership for all students (Texas Education Agency, 2012). Although attendance is the not an indicator for Texas high schools in meeting AYP, it is included as a variable in this study because all schools are required to report their attendance rates to the Texas Education Agency and the student group variables are based upon reported participation rates for each assessment.

Additionally, previous literature indicates that attendance is crucial in student success (Drewry, Burge, & Driscoll, 2010; M. Dynarski & Gleason, 2002; Steward, Devine Steward, Blair, Jo, & Hill, 2008; Stout & Christenson, 2009; Texas Education Agency, 2012).

Meeting these criterion is exceedingly difficult for campuses, particularly those located in urban areas. The ramifications are costly if a school fails to meet AYP for two years in a row. Regardless of the indicator in which they fail to meet the standard, students in the failing school must be given an option to transfer to another public school that has not been identified as needing improvement, with transportation provided by the district. The district must also provide students opportunities for supplemental services, such as tutoring provided by outside agencies, remedial classes, and summer school. In addition, if the school fails to meet AYP for a third consecutive year, the district must take corrective action. During the fourth year of AYP failure, the district must develop plans for restructuring the school, and if the school fails in the fifth year, the district is required to implement the restructuring, which may include being converted into a charter school, or being closed (Lee & Reeves, 2012; Linn, 2003; "NCLB," 2002). Thus, the consequences for failing to meet AYP are high. These consequences all have potentially high financial costs for the district, which are unfunded by the federal government.

*Framework of the present study:*

Thus, the aim of the present study is to analyze when campuses are most at risk for failing to meet AYP and which variables are most influential in the failure to meet AYP. This study will provide a framework for future research to broaden the scope of analysis from Texas senior high schools to an analysis of both public high schools and charter schools to determine whether there is a difference in which schools fail to meet AYP, when they fail to meet AYP, and which variables are most significant in that failure to meet AYP. Understanding where and when successes and failures occur and why would be beneficial to campus and district leaders to allow them to more effectively address the needs of students. Additionally, future researchers could expand this study to other states and, potentially, a national study. Hence, there is one central research question for this study and nine supporting questions.

The central research question is:

*To what extent is the probability of Texas public and charter high schools failing to meet Adequate Yearly Progress (AYP) for the first time under NCLB from 2003-2011 associated with the multiple variables that are known to be associated with school performance and AYP regulations?*

## METHODS:

***Sample***

This study is a secondary analysis of publically available data from the Texas Education Agency (TEA) Division of Performance Reporting, Adequate Yearly Progress (TEA, 2012) and the National Center for Education Statistics (NCES) Common Core of Data (CCD) (NCES, 2006) . The TEA dataset is a comprehensive state-wide dataset that provides AYP results for all eligible schools in the State of Texas, as well as all variables associated with the AYP results, while the NCES CCD annually collects fiscal and non-fiscal public school data for all schools in the U.S. The sample for this study includes the entire population of schools identified as senior high schools in the state of Texas (*n*=1721). Utilizing these databases provides a unique opportunity to analyze the impact of NCLB on a large and diverse population, as well as the factors that contribute to success or failure of schools meeting AYP. Although both datasets provide information for all subpopulations that are included in determining AYP, some data is masked in order to adhere to the requirements of the Family Educational Rights and Privacy Act (FERPA), for schools with very low subpopulations.

*Variables Included in the Analytic Model*

The dependent variable in the following analysis is the probability of a high school failing Adequate Yearly Progress (AYP) status for the first time between the years 2003 and 2011 in Texas. AYP is the term used by NCLB to determine whether schools are making adequate improvements in educating students. AYP requires all public schools to meet criteria based upon reading and math assessments, student groups, and another indicator. In Texas, the other indicator for high schools it is graduation rate. Student groups include Black, Hispanic, White, English language learners, economically disadvantaged, and special education students.

We used previous literature to guide the selection of covariates and controls for this study (Barton, 2006; Bridgeland, Dilulio, & Morison, 2006; Drewry et al., 2010; M. G. P. Dynarski, 2002; Elledge, Le Floch, Taylor, & Anderson, 2009a; Hursh, 2005; National Research Council, 2001; Steward, Steward, Blair, Jo, & Hill, 2008; Stout & Christenson, 2009; Warren & Edwards, 2005). Utilizing the recommendations of previous research (Bowers & Lee, 2013), district urbanicity categories, based upon the U.S. Census metrocentric codes, were derived from the NCES CCD (NCES, 2006) and converted into the variables *Urban*, *Small Town*, and *Rural*, with Suburban as the reference group. Background and control variables include *Percent African American Students, Percent Hispanic Students, Percent Asian Students, Percent Students on Free and Reduced Lunch, Teacher Pupil Ratio, Total Enrollment, Attendance, Graduation, and Percent Met Standard on*

**TABLE 1:** *Descriptive Variables and Frequencies by Campus*

| **Name** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|
| Urban | 0.287 | 0.452 | 0 | 1 |
| Town | 0.130 | 0.336 | 0 | 1 |
| Rural | 0.316 | 0.465 | 0 | 1 |
| Attendance | 92.681 | 0.035 | 48.10 | 100 |
| Graduation | 82.238 | 18.824 | 0 | 100 |
| Math | 63.294 | 17.901 | 2 | 99 |
| Percent Free & Reduced Lunch students | 0.381 | 0.218 | 0 | 0.99 |
| Total Student Enrollment | 767.680 | 867.320 | 0 | 4679 |
| Pupil Teacher Ratio | 13.080 | 6.868 | 0 | 201.00 |
| Percent African American students | 0.121 | 0.172 | 0 | 1.00 |
| Percent Hispanic students | 0.361 | 0.310 | 0 | 1.00 |
| Percent Asian students | 0.015 | 0.035 | 0 | 0.39 |
| *N* | 1721 | | | |

*TAKS Math Scores in 10th Grade.* Table 1 provides descriptive statistics for the variables, including mean, standard deviation, minimum, and maximum.

*Analysis*

For all models in this study, we used discrete time hazard modeling to model the probability of a high school in the state of Texas failing to meet AYP for the first time between 2003 and 2011. As a subset of survival analysis (Singer & Willett, 2003), discrete time hazard modeling is an attractive modeling framework for this type of data, since the event under consideration here is a heavily dependent and conditional event that is nested within time, a feature of the data which we aim to take advantage of through the use of discrete time hazard modeling, following the recommendations of past education research that has used this method, particularly examining the Texas context (Bowers, 2010; Bowers & Lee, 2013; Bowers & Metzger, 2010b; Bowers, Metzger, & Militello, 2010a). Briefly, the data are dependent and conditional due to the point that once a school has failed AYP for the first time, they cannot fail AYP for the first time again, which violates one of the central assumptions of inferential statistics and is accounted for through discrete time hazard modeling and the use of a "unit-period" dataset (Singer & Willet, 2003) in which each unit (here the entire population of secondary schools in Texas) are represented in the dataset multiple times, once for each year of data. Time varying and time invariant variables can then be estimated through a modified logistic regression framework in which each parameter is estimated on the probability of the "hazard" of the event occurring within any one year, controlling for the point that the event could have already happened in a previous year, and so the overall sample size and conditional probabilities for each year must be correctly adjusted and accounted for (Singer & Willett, 2003).

Following the methods recommended for discrete time hazard modeling (Singer & Willett, 2003), the campus-level dataset was converted into a school-period dataset, in which each school was represented in the dataset with nine rows of data, one for each year, with event defined as AYP status the first time a campus failed to meet AYP. Campuses who never failed to meet AYP were censored at the end of the study in 2011. New campuses that were founded after 2003 were included in the study with missing data for all variables up until the time of entry and the first opportunity to fail to meet AYP.

Thus, the dependent variable in all models was the probability of a campus failing AYP for the first time within any one year, 2003 through 2011. Following the recommendations from the previous literature, we included the above noted variables as either time dependent or time invariant variables in the subsequent models. Time invariant variables included in the study were the locale variables of Urban, Town and Rural. Time variant variables include percent attendance, percent graduation, percent met standard on mathematics TAKS in 10th grade, percent African American students, percent Hispanic students, percent Asian students, percent students on free and reduced lunch, teacher pupil ratio, and total enrollment.

Using the nomenclature recommended by Singer & Willett (2003) for these types of models, logit discrete-time hazard models were estimated and fit to the AYP campus level data, as well as the calculation of life tables, and estimated hazard and survival functions. The full model used for this analysis can be written as:

$$logit(Y) = \alpha_1 X_{period}^{1} + \alpha_2 X_{period}^{2} + \alpha_3 X_{period}^{3} + \alpha_4 X_{period}^{4} + \alpha_5 X_{period}^{5} + \alpha_6 X_{period}^{6} + \alpha_7 X_{period}^{7} + \alpha_8 X_{period}^{8} + \alpha_9 X_{period}^{9} + \beta_{URBAN} \text{ X } Urban + \beta_{TOWN} \text{ X } Town + \beta_{RURAL} \text{ X } Rural + \beta_{ATTENDANCE} \text{ X } Attendance + \beta_{GRADUATION} \text{ X } Graduation + \beta_{MATH} \text{ X } Math + \beta_{FREE \, \& \, REDUCED \, LUNCH} \text{ X } Free \, \& \, Reduced \, Lunch + \beta_{TOTAL \, ENROLLMENT} \text{ X } Total \, Enrollment + \beta_{PUPIL \, TEACHER \, RATIO} \text{ X } Pupil \, Teacher \, Ratio + \beta_{AFRICAN \, AMERICAN} \text{ X } African \, American + \beta_{HISPANIC} \text{ X } Hispanic + \beta_{ASIAN} \text{ X } Asian.$$

## RESULTS

The purpose of this study is to examine when campuses are most at risk for failing to meet AYP and which variables are most associated with campus AYP failure. Adequate Yearly Progress data and variables included in the determination of meeting AYP were recorded for the entire population of schools identified as senior high schools in the state of Texas (TEA, 2012). The overall sample, *n*=1721, included all senior high school campuses in existence at any time between 2003-2011. Locale-urbanicity variables were downloaded from NCES CCD (NCES, 2006). Because campuses are not required to meet AYP their first year, that data is recorded as missing for schools that were founded during this time period. Additionally, although both datasets provide information for all subpopulations that are included in determining AYP, some data is masked in order to adhere to the requirements of the Family Educational Rights and Privacy Act (FERPA), for schools with very low subpopulations. All masked data is also recorded as missing.

In this section, we first present a description of the overall AYP results in the state of Texas, to provide background and justification for the study. Second, we discuss the hazard probability to show the proportion of campuses in the sample at risk for not meeting AYP. Third, we detail the fit of the full model to the data and then finally, turn to a discussion of the results.

*Examining Overall State of Texas AYP Results*

As has been suggested in longitudinal data analysis literature (Mills, 2011; Singer & Willett, 2003), survival analysis was utilized to investigate the event occurrence of campuses failing to meet AYP. This type of analysis has been shown to be superior to simple means and weighted means when analyzing the risk of a terminal event (Singer & Willet, 1995, 2003), such as failing to meet AYP, in which a campus, once it has failed AYP, cannot reverse its status and become a school that has never failed to meet AYP. Survival analysis allows us to examine the campuses during each school year that are still at risk for not meeting AYP, rather than aggregating all years in the dataset. This requires censoring, or removal of, two types of campuses from calculations during each school year – those that failed to meet AYP, since they cannot experience the event again for the first time, and those that ceased to be in existence. Table 2 presents a life table, as suggested by Singer and Willett (2003), detailing the AYP event histories for each year of the sample of 1721 campuses. The life table, presented with hazard and survival estimates, is a more realistic representation over time than previous methods.

The estimated hazard probability shows the proportion of campuses in the sample at the end of each school year still at risk of failing to meet AYP who failed to meet AYP for that school year. In other words, any campus that missed AYP at the end of the school year were thus categorized as failing to meet AYP *(Table 2, fifth column; Figure 1)*. The 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 campuses experiencing the event across time periods while controlling for the campuses that had experienced the event and are not in the dataset for subsequent years. For example, a hazard probability of .144 in the fifth column of Table 2 for the year 2005, and plotted in Figure 1 (Panel A, right) as YR2005, indicates that for individual campuses, 14.4% of them experienced the event of missing AYP in 2005 in Texas.

Using these life table calculations, estimating the hazard function for failing to meet AYP, one can see that the probability of failing to meet AYP for this dataset was highest during the first year of AYP in the dataset in 2003, with 30.9% of campuses experiencing the event. Probabilities drop in 2004 7%, rises again in 2005 to 14.4% and then subsides until 2011 when 19.4% of the campuses

**Table 2:** *Life table describing the number of campuses meeting AYP, the hazard and survival probabilities of meeting AYP, and the median lifetime of campuses meeting AYP*.

| Year | Number | | | Proportion | |
|---|---|---|---|---|---|
| | Campuses meeting AYP at the beginning of the year | Campuses who failed to meet AYP at the end of the year | Campuses censored at the end of the year | Hazard probability of campuses not meeting AYP | Survival probability of campuses meeting AYP |
| 2003 | 1219 | 377 | 0 | 0.309 | 0.690 |
| 2004 | 842 | 59 | 0 | 0.070 | 0.642 |
| 2005 | 783 | 113 | 0 | 0.144 | 0.549 |
| 2006 | 670 | 16 | 0 | 0.023 | 0.536 |
| 2007 | 654 | 41 | 0 | 0.062 | 0.502 |
| 2008 | 613 | 53 | 0 | 0.086 | 0.459 |
| 2009 | 560 | 5 | 0 | 0.008 | 0.455 |
| 2010 | 555 | 4 | 0 | 0.007 | 0.452 |
| 2011 | 551 | 107 | 444 | 0.194 | 0.364 |
| Total | 1721 | 775 | | | |

experienced the event of not meeting AYP. Thus, the most hazardous years for failing to meet AYP were 2003, 2005, and 2011.

The final far-right column of Table 2 indicates the survival function which presents the data in a cumulative manner, displaying the data points as the percentage of the full sample that "survived", i.e. those that did not experience the event of failing to meet AYP, while controlling for campuses that had already experienced the event and dropped out of the dataset (*Table 3, sixth column, Figure 2)*. At the end of 2003, 69% of campuses had not experienced the event of failing to meet AYP. The number of campuses who had not experienced the event consistently dropped during each year of the study, with only 36.4% of campuses surviving the event by 2011. Figure 1, panel A, plots these overall hazard and survival probabilities as a function through years 2003-2011.

While the calculation of overall AYP results using survival analysis and life tables are interesting, these numbers do not give an indication of which variables are most associated with failure to meet AYP. The main focus of this study is to combine and analysis of the timing of failing to meet AYP with an analysis of the variables required by NCLB to meet AYP, controlling for the demographics and locale of the school. As discussed above, campuses must not only meet AYP criteria for the overall campus but also for the subpopulations and another indicator, which is graduation in Texas. Additionally, as discussed above, urban areas and areas with densely populated groups within the subpopulations have a difficult time meeting the requirements of AYP. For many of these variables, the hazard and survival functions vary across the state when examining these variables individually. As an example, the locale urbanicity categories for the high schools were used to disaggregate the hazard and survival functions, plotted in Figure 1 panel B. As noted in Figure 1, urban schools in 2003-2005 experienced the highest risk of failing to meeting AYP in the state, followed by suburban schools, small towns, and rural schools. However, by 2006, many of the schools most at risk of failing AYP had done so, and so the

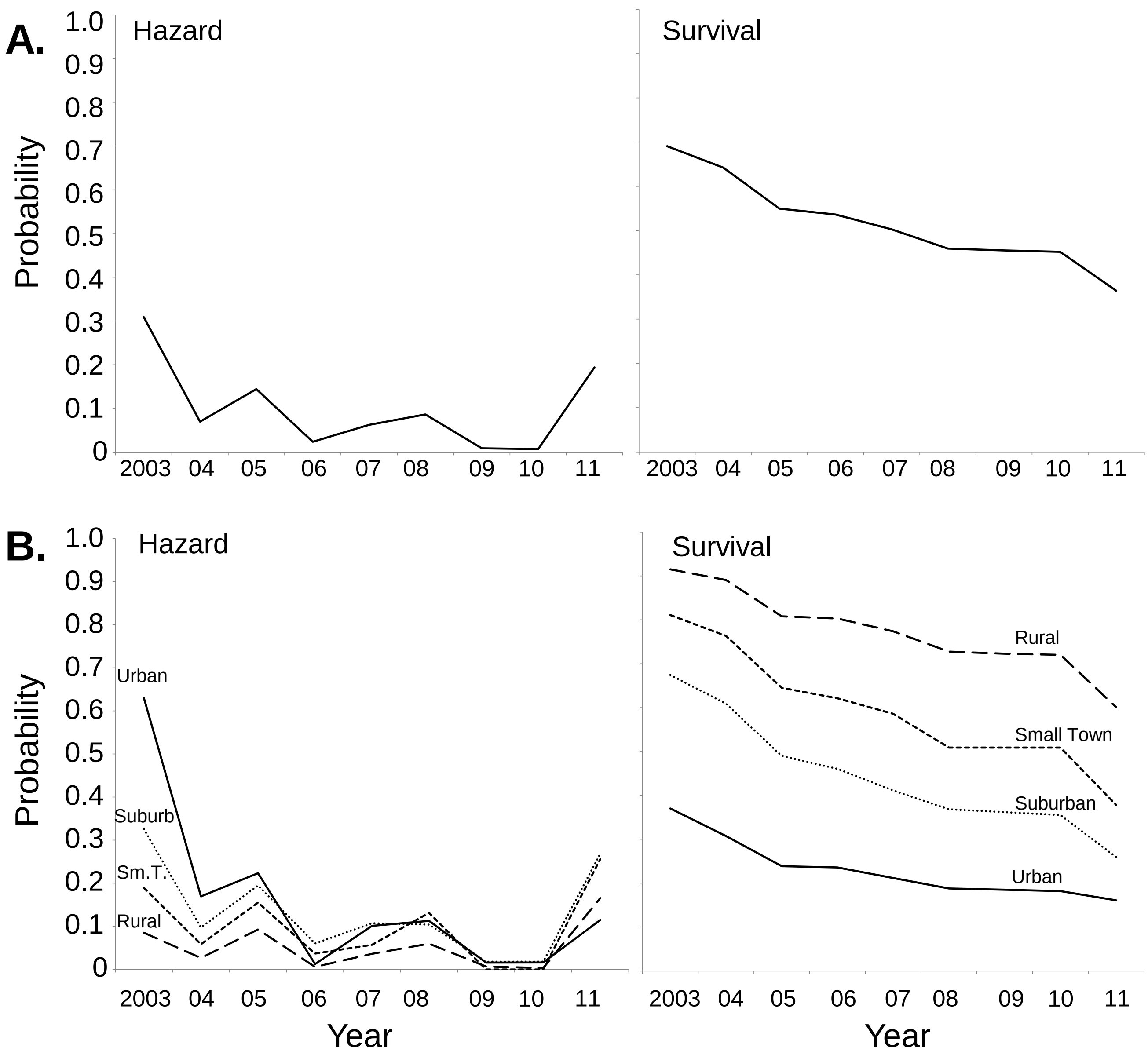


**Figure 1**: Estimated hazard and survivor functions for failing AYP for the first time from 2003 through 2011 for all secondary schools in Texas. Panel A plots the estimated hazard and survivor functions for the full sample. The two greatest years of the hazard of failing AYP for the first time were 2003 and 2011, with 2009 and 2010 being the least hazardous years. The survival function shows that by 2011, only 36.4% of all secondary schools in Texas had not failed AYP at least once over the nine years. Panel B plots the estimated hazard and survivor functions disaggregated by school locale: urban, suburban, small town and rural. Note that after the first three years in the dataset, hazard rates converge as the schools that were most likely to fail AYP had most likely done so by 2006 for the first time within each of the different contexts.

hazard function appropriately accounts for this. Year 2011 also of interest as the hazard of failing AYP for the first rises substantially. The corresponding survival functions show the cumulative survival of each of the different school locals over time, demonstrating that rural schools were the least likely to fail AYP throughout the full time period, followed by small towns, and suburban school, while urban schools experienced the highest risk of hazard generally throughout the dataset. Thus, these types of variables individually appear to covary in interesting ways through time with school AYP failure. We turn next to examining the variables together in a full discrete time hazard model.

To address the question of the extent to which different variables, including graduation, attendance, location, and subpopulations may correspond to a campus's increased risk of failing to meet AYP, we used discrete-time hazard models within a logistic regression framework (see methods). As has been previously argued, a discrete-time hazard model using logit regression is superior for predicting a campus's risk for failing to meet AYP. This is because this type of modeling framework appropriately handles the dichotomous outcome variable (meeting AYP or not for the first time within a year), as well as campuses ceasing to exist through censoring or being founded across time, and the ability to include the effect of time within the equation. Rather than experiencing a continuous change over time, campuses experience AYP failure at one point in the year with open periods between each discrete period. Thus, a discrete-time hazard model is appropriate for modeling the risk of failing to meet AYP and testing the different variables for the ability to predict which campuses may fail to meet AYP.

A discrete-time hazard model was fitted to the data by estimating parameters for each time period and for each of the variables using logistic regression. 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 variable collected on a sample are then fit to the model as β estimates, and then the model fit is assessed. Four discrete-time hazard models are presented in Table 3, listing parameter estimates and significance levels, standard errors for each estimate (in parentheses), the overall *N* of the campus-period dataset at each time-point, and tests of model goodness of fit, including -2Log-likelihood, Chi-Square, and Cox & Snell pseudo $R^2$.

The model was tested in a forward stepwise fashion, starting first with the "unconditional" model testing only the nine intercepts for time through the years. Estimated parameter coefficients are reported in logits. Time alone in the model explains an estimated 53.2% of the variance in the probability of a school failing to make AYP in any one year from 2003-2011. However, as noted in the literature (Singer & Willett, 2003), variance explained calculations of logistic regressions are estimates only, and must be interpreted cautiously since there are currently no precise means to calculate a true amount of variance explained within this modeling framework. We present the Cox & Snell pseudo-R-squared as a means to help assess fit and the subsequent models. As noted in Table 3 Model A as well as Table 2 and Figure 1, the most hazardous years of failing to make AYP in the unconditional model were 2003 and 2011. Continuous variables included in the model were all standardized (z-scored). Model B includes the locale variables, while Model C adds the school background and demographic control variables. As our final model, Model D provides the full parameter estimates, as well as the odds as an estimate of effect size for each of the variables included in the model. As noted in Table 3, other than time, multiple background and control variables were significant in predicting the probability of failing AYP. First, while urban and small town were not statistically significant, Rural schools were significantly less likely to fail AYP than suburban schools, controlling for the other variables in the model. As noted by Singer & Willett (2003), odds less than 1.00 can be difficult to interpret, thus they recommend inverting the number to aid in interpretation. Thus, Rural schools were 1.427 times less likely to fail AYP than Suburban schools in the model. The percent student variables measuring ethnicity were all significant in the model, such that schools with a one standard deviation higher percentage of African American students were 1.522 times more likely to fail AYP than schools at the average. Interpretation of logistic regression coefficients beyond a difference of 0 to 1 though are problematic due to the non-linear relationship, so we stress that interpretations such as this be interpreted with caution.

Additionally, schools with a one standard deviation higher pupil teacher ratio were 2.183 times more likely to fail AYP, and schools with a one standard deviation higher enrollment were 2.388 times more likely. As predicted since it is part of the calculation for AYP failure, percent met standard in mathematics on the state standardized TAKS test was negative and significant in the model. As the largest effect in the model, inverting the odds of 0.404 gives an odds of schools with a one standard deviation higher average percent met standard failing AYP in any one year 2.475 times less than schools, all other variables being equal. Schools with higher percent attendance were 1.318 times less likely to fail AYP. Interestingly, graduation rate was not significant in the final model. Overall, the final model accounted for an estimated 61.1% of the variance in the probability of a school failing AYP in Texas between 2003-2011.

**Table 3:** *Results of fitting four discrete-time hazard models to the year of failing to meet AYP using parameter estimates.*

| | *Model A* | *Model B* | *Model C* | *Model D* | *Odds* |
|---|---|---|---|---|---|
| Parameter Estimates and Asymptotic Standard Errors | | | | | |
| D2003 | -.804*** (.062) | -.641*** (.092) | -.841*** (.108) | -.567*** (.133) | 0.567 |
| D2004 | -2.586*** (.135) | -2.301*** (.151) | -2.389*** (.165) | -2.805*** (.207) | 0.060 |
| D2005 | -1.780*** (.102) | -1.422*** (.122) | -1.285*** (.133) | -1.456*** (.159) | 0.233 |
| D2006 | -3.711*** (.253) | -3.370*** (.262) | -3.228*** (.273) | -3.566*** (.301) | 0.028 |
| D2007 | -2.705*** (.161) | -2.339*** (.176) | -2.079*** (.186) | -2.114*** (.210) | 0.121 |
| D2008 | -2.358*** (.144) | -1.959*** (.160) | -1.624*** (.172) | -1.806*** (.204) | 0.164 |
| D2009 | -4.710*** (.449) | -4.335*** (.455) | -3.976*** (.459) | -4.173*** (.477) | 0.015 |
| D2010 | -4.925*** (.502) | -4.549*** (.507) | -4.163*** (.511) | -4.506*** (.537) | 0.011 |
| D2011 | -1.423*** (.108) | -.949*** (.131) | -.501*** (.143) | | |
| Urban | | .645*** (.110) | .104 (.136) | .027 (.168) | |
| Town | | -.425*** (.132) | -.259 (.147) | -.177 (.178) | |
| Rural | | -1.107*** (.113) | -.624*** (.132) | -.355* (.165) | 0.701 |
| Z Percent Free & Reduced Price Lunch | | | .285*** (.074) | .040 (.101) | |
| Z Percent African American students | | | .332*** (.064) | .420*** (.086) | 1.522 |
| Z Percent Hispanic students | | | .358*** (.077) | .409*** (.100) | 1.505 |
| Z Percent Asian students | | | -.305*** (.057) | -.281*** (.071) | 0.755 |
| Z Pupil Teacher ratio | | | .297*** (.087) | .781*** (.142) | 2.183 |
| Z Total school enrollment | | | .760*** (.059) | .871 (.083) | 2.388 |
| Z Percent Met Standard Mathematics | | | | -.907*** (.103) | 0.404 |
| Z Percent Attendance | | | | -.276* (.089) | 0.759 |
| Z Graduation rate | | | | .039 (.078) | |
| *N* | 6447 | 6447 | 6124 | 4975 | |
| n parameters | 8 | 12 | 18 | 20 | |
| -2Log-likelihood | 4047.106 | 3785.436 | 3338.522 | 2199.564 | |
| Chi-Square | 4890.33 | 5152.004 | 5151.145 | 4697.251 | |
| Cox & Snell pseudo-$R^2$ | 0.532 | 0.550 | 0.569 | 0.611 | |

## DISCUSSION:

Overall, this study came to three main findings. First, as one of the first studies to use a discrete time modeling framework to examine the "hazard" of a school failing to meet AYP, this study shows that this type of modeling framework works well in this policy domain and can provide interesting findings to help examine not only the most hazardous years for schools, but what variables are most associated with a school failing AYP. This is important since under NCLB and Texas education policy, failing AYP activates a range of sanctions that most schools and school staff would wish to avoid. Second, this study finds that variables that not part of the AYP calculation policy are significantly related to the probability of a school failing AYP in any one year. These variables include school locale, with Rural schools being much less likely to fail AYP than Suburban schools, pupil teacher ratio and school enrollment. We note that from a policy perspective, these findings suggest that the AYP policy may be affecting schools differently in different contexts, which would appear to go against the intention of the policy to sanction and reward schools purely on a performance basis across their subgroups. Third, multiple variables included within the AYP calculation were significant in the model. However, we were surprised that the graduation rate was not significant in the model. This may be due to multicolinearity issues within the model, however our preliminary descriptive analysis suggested that this was not an issue (data not shown). We suggest further research in this area.

*Limitations*

While we argue that our model is robust, this study is limited in the following ways. First, the beginning of time in the model was year 2003. While this corresponds with the beginning of the application of the full effects of NCLB in Texas, as the model state for NCLB, Texas' accountability policies had been in effect for a decade previously. Thus, the true "beginning of time" for a hazard model would be better defined at an earlier time point. However, as a second issue, prior to 2003, the state standardized test was the TAAS, rather than TAKS, and so not only was the accountability regime somewhat different, estimating a model with two different tests through time is problematic. Thus, we selected the first year of the study as 2003. Third, within the discrete time hazard modeling framework, when hazard rates fall over time, unobserved heterogeneity can be a issue (Singer & Willet, 2003) as all time conditional models, such as the one used here, have an assumption of no unobserved heterogeneity – in that the hazard function for any school in the dataset is dependent only on that school's predictors. We acknowledge that unobserved heterogeneity could be an explanation for the hazard functions noted here and we encourage future work in this area. However, as one of the first studies to examine the probability of failing AYP across an entire population of all high schools in a large state using a method that appropriately controls for the conditional dependent nature of the dataset, we argue that this study lays the foundation for future studies that could examine if there are unobserved variables that are contributing to the overall risk.

## Recommended Citation Format:

Pruitt, P.L.,Bowers, A.J. (2014) At What Point Do Schools Fail to Meet Adequate Yearly Progress and What Factors are Most Closely Associated with Their Failure? A Survival Model Analysis. A paper presented at the annual meeting of the Association of Education Finance and Policy (AEFP), San Antonio, TX: March 2014.

## REFERENCES:


Amrein, A. T., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32.

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing high schools? *American Educational Research Journal, 44*(3), 559-593. doi: 10.3102/0002831207306768

Barton, P. E. (2006). The dropout problem: Losing ground. *Educational Leadership, 63*(5), 14-18.

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading: Addison-Wesley.

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: doi:10.1080/00220670903382970

Bowers, A. J., & Lee, J. (2013). Carried or defeated? Examining the factors associated with passing school district bond elections in Texas, 1997-2009. . *Educational Administration Quarterly, 49*(5), 732-767. doi: 10.1177/0013161x13486278

Bowers, A. J., & Metzger, S. A. (2010b). Knowing what matters: An expanded study of school bond elections in Michigan, 1998-2006. *Journal of Education Finance, 35*(4), 374-396. doi: 10.1353/jef.0.0024

Bowers, A. J., Metzger, S. A., & Militello, M. (2010a). Knowing the odds: Parameters that predict passing or failing school district bonds. *Educational Policy, 24*(2), 398-420. doi: 10.1177/0895904808330169

Bracey, G. W. (1996). International comparisons and the condition of American education. *Educational Researcher, 25*(1), 5-11. doi: 10.3102/0013189x025001005

Bridgeland, J. M., Dilulio, J. J. J., & Morison, K. B. (2006). The silent epidemic: Perspectives of high school dropouts (pp. 1-35). Washington D.C. : Civic Enterprises.

Bulkley, K., & Fisler, J. (2003). A decade of charter schools: From theory to practice. *Educational Policy, 17*(3), 317-342. doi: 10.1177/0895904803017003002

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental educational services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis, 29*(2), 115-133. doi: 10.3102/0162373707302035

California Department of Education. (2012). *Standardized testing and reporting program: Annual report to the legislature* Sacramento: California Department of Education.

Carroll, S., & Carroll, D. (2002). *Statistics made simple for school leaders*. Lanham: Scarecrow Education.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101. doi: 10.1111/1467-8721.ep10768783

DeBray-Pelot, E. H. (2007). NCLB's transfer policy and court-ordered desegregation: The conflict between two federal mandates in Richmond County, Georgia, and Pinellas County, Florida. *Educational Policy, 21*(5), 717-746. doi: 10.1177/0895904806289220

DeBray-Pelot, E. H., & McGuinn, P. (2009). The new politics of education: Analyzing the federal education policy landscape in the post-NCLB era. *Educational Policy, 23*(1), 15-42. doi: 10.1177/0895904808328524

Dee, T. S., & Jacob, B. A. (2010). The impact of No Child Left Behind on students, teachers, and schools. *Brookings Papers on Economic Activity* (pp. 149-207). Washington D.C.

Dee, T. S., & Jacob, B. A. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446. doi: 10.2307/23018959

Dee, T. S., Jacob, B. A., M., a., & Ladd, H. F. (2010). The impact of No Child Left Behind on students, teachers, and schools *Brookings Papers on Economic Activity*, 149-207. doi: 10.2307/41012846

DiMartino, C., & Scott, J. (2013). Private sector contracting and democratic accountability. *Educational Policy, 27*(2), 307-333. doi: 10.1177/0895904812465117

Drewry, J. A., Burge, P. L., & Driscoll, L. G. (2010). A tripartite perspective of social capital and its access by high school dropouts. *Education and Urban Society, 42*(5), 499-521. doi: 10.1177/0013124510366799

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, 7*(1), 43-69.

Dynarski, M. G. P. (2002). How can we help? What we have learned from recent federal dropout prevention evaluations. *Journal of Education for Students Placed at Risk, 7*(1), 43-69.

Elledge, A., Le Floch, K. C., Taylor, J., & Anderson, L. (2009a). State and local implementation of the No Child Left Behind Act *Volume V -- Implementation of the 1 percent rule and 2 percent interim policy options* (Vol. V, pp. 1-105). Washington D. C.: American Institutes for Research.

Elledge, A., Le Floch, K. C., Taylor, J., & Anderson, L. (2009b). State and local implementation of the No Child Left Behind Act: Volume V—Implementation of the 1 percent rule and 2 percent interim policy options (Vol. V, pp. 1-105). Washington D. C.: American Institutes for Research.

Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher, 36*(5), 268-278. doi: 10.3102/0013189x07306556

Fuller, E. J., & Johnson, J. F. (2001). Can state accountability systems drive improvements in school performance for children of color and children from low-income homes? *Education and Urban Society, 33*(3), 260-283. doi: 10.1177/0013124501333003

Fusarelli, L. D. (2004). The potential impact of the No Child Left Behind Act on Equity and diversity in American education. *Educational Policy, 18*(1), 71-94. doi: 10.1177/0895904803260025

Haertel, E. H. (1999). Validity Arguments for High-stakes Testing: In Search of the Evidence. *Educational Measurement: Issues and Practice, 18*(4), 5-9.

Hannaway, J., McKay, S., & Nakib, Y. (2002). *Reform and resource allocation: National trends and state policies*. Washington D. C.: National Center for Education Statistics.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management, 24*(2), 297-327. doi: 10.2307/3326211

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis, 32*(2), 273-298. doi: 10.3102/0162373710361640

Henig, J. R. (1990). Choice in public schools: An analysis of transfer requests among magnet schools. *Social Science Quarterly (University of Texas Press), 71*(1), 69-82.

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal, 31*(5), 605-622.

Hursh, D. (2007). Assessing 'No Child Left Behind' and the rise of neoliberal education policies. *American Educational Research Journal, 44*(3), 493-518.

Johnson, D. P. (2007). Challenges to No Child Left Behind: Title I and Hispanic students locked away for later. *Education and Urban Society, 39*(3), 382-398. doi: 10.1177/0013124506297962

Jozwiak, K., & Moerbeek, M. (2011). Power analysis for trials with discrete-time survival endpoints. *Journal of*

*Educational and Behavioral Statistics*. doi: 10.3102/1076998611424876
Koyama, J. P. (2012). Making failure matter: Enacting No Child Left Behind's standards, accountabilities, and classifications. *Educational Policy, 26*(6), 870-891. doi: 10.1177/0895904811417592
Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208. doi: 10.3102/0162373711429989
Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231. doi: 10.3102/0162373711431604
Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3-13. doi: 10.3102/0013189x032007003
Loveless, T. (2006). The peculiar politics of No Child Left Behind (pp. 1-38). Washington D.C.: The Brookings Institute: Brown Center on Education Policy.
May, J. J. (2006). The charter school allure: Can traditional schools measure up? *Education and Urban Society, 39*(1), 19-45. doi: 10.1177/0013124506291786
McDill, E. L., Natriello, G., & Aaron, M. P. (1986). A population at risk: Potential consequences of tougher school standards for student dropouts. *American Journal of Education, 94*(2), 135-181.
Mills, M. (2011). *Introducing survival and event history analysis*. Thousand Oaks: SAGE Publications, Inc.
Mintrop, H., & Sunderman, G. L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement—And why we may retain it anyway. *Educational Researcher, 38*(5), 353-364. doi: 10.3102/0013189x09339055
National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing*. In Division of Behavioral and Social Sciences and Education (Series Ed.) A. Beatty, U. Neisser, W. Trent & J. P. Heubert (Eds.), (pp. 67). Retrieved from http://www.nap.edu/catalog/10166.html
NCES. (2006). Common Core of Data. from https://nces.ed.gov/ccd/
No Child Left Behind Act of 2001, 20 U.S.C. §6301, Pub. L. No. 107-110 1-670 (2002).
Novak, J. R., & Fuller, B. (2003). Penalizing diverse schools? PACE policy brief 03-04 (pp. 12). Berkeley: Policy Analysis for California Education.
Orfield, G., & Lee, C. (2005). Why segregation matters: Poverty and Educational Inequality *The Civil Rights Project* (pp. 1-47). Cambridge: Harvard University.
Price, H. E. (2010). Does No Child Left Behind really capture school quality? Evidence From an urban school district. *Educational Policy, 24*(5), 779-814. doi: 10.1177/0895904810376564
Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure*. Washington D.C.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press, Inc. .
Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory into Practice, 42*(1), 12-17.
Steward, R. J., Devine Steward, A., Blair, J., Jo, H., & Hill, M. F. (2008). School attendance revisited: A study of urban African American students' grade point averages and coping strategies. *Urban Education, 43*(5), 519-536. doi: 10.1177/0042085907311807
Steward, R. J., Steward, A. D., Blair, J., Jo, H., & Hill, M. F. (2008). School attendance revisited: A study of urban african american students' grade point averages and coping strategies. *Urban Education, 43*(5), 519-536. doi: 10.1177/0042085907311807
Stiefel, L., Schwartz, A. E., & Chellman, C. C. (2007). So many children left behind: Segregation and the impact of subgroup reporting in No Child Left Behind on the racial test score gap. *Educational Policy, 21*(3), 527-550. doi: 10.1177/0895904806289207
Stout, K. E., & Christenson, S. L. (2009). Staying on track for high school graduation. *The Prevention Researcher, 16*(3).
TEA. (2012). Adequate Yearly Progress. from http://www.tea.state.tx.us/ayp/
Texas Education Agency. (2010). House Bill 3 transition plan (pp. 433). Austin: Texas Education Agency.
Texas Education Agency. (2012). *2012 Adequate yearly progress (AYP) guide for Texas public school districts and campuses*. Austin: Texas Education Agency.
U.S. Department of Education, & The National Commission on Excellence in Education. (1983) A nation at risk: The imperative for educational reform. Retrieved from https://www2.ed.gov/pubs/NatAtRisk/intro.html.
No Child Left Behind Act of 2001, 20 U.S.C., Pub. L. No. 107-110 1-670 §6301 Stat. (2002).
Vasquez Heilig , J., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110. doi: 10.3102/0162373708317689
Warren, J. R., & Edwards, M. R. (2005). High school exit examinations and high school completion: Evidence from the early 1990s. *Educational Evaluation and Policy Analysis, 27*(1), 53-74.
Weckstein, P. (2003). Accountability and student mobility under Title I of the No Child Left Behind Act. *The Journal of Negro Education, 72*(1), 117-125. doi: 10.2307/3211295

Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy*, *26*(2), 268-308. doi: 10.1177/0895904810386588