HERC Houston Education Research Consortium | RICE | KINDER INSTITUTE FOR URBAN RESEARCH

RESEARCH BRIEF

for the Houston Independent School District

Volume 3, Issue 5 – September, 2015

Evaluating High School Dropout Indicators and Assessing Their Strength

By D. Diego Torres, Ph.D., Amanda Bancroft, M.A., and Kori Stroub, M.A.[1]

Using administrative student- and school-level data from the Houston Independent School District, this study assessed both the accuracy and relative strength of a number of predictors of dropping out of high school, with the aim of providing policymakers with the best information to address the dropout crisis. We find that, while most of the indicators used in this study are better predictors of high school dropout than not, the strongest by far were being over age for grade at the start of ninth grade, receipt of an F during any grading period during eighth grade, having had a disciplinary issue in eighth grade, and failure to meet the eighth grade math standard. These results are discussed and recommendations for prevention/intervention are given.

BACKGROUND

Finding accurate predictors of high school non-completion continues to rank as a major concern among policymakers and social science scholars whose research focuses on inequality in educational attainment (see Rumberger & Lim, 2008, for a review of the literature). A primary reason for this is that high school dropout is linked to poor outcomes in late adolescence and adulthood, including unemployment, crime, and incarceration (Anderson, 2014; Arum & LaFree, 2008; Rumberger & Lamb, 2003). The greatest relationship between high school dropout and these factors are found among racial and ethnic minorities, the socioeconomically disadvantaged, and males. These outcomes reproduce for another generation the structural inequality believed to be the cause of early school leaving in the first place.

Determining accurately who is likely to drop out is most important because schools and districts aim to implement interventions to prevent them from doing so (Gleason & Dynarski, 2004). Since they are ultimately limited in the amount of resources they have at their disposal to solve problems stemming from inequality, districts seek to allocate funds in a way that maximizes efficacy. That is, they do not want to target an intervention at those who do not need it, and, on the other hand, they do not wish to deny an intervention to those who do need it. Bowers, Sprott, and Taff (2013) highlight a shortcoming of the dropout literature, namely that many studies report the overall number of students having a specific dropout flag while then going on to report inferential statistics using logistic regression. ***Few studies report measures of accuracy, of which the main components consist of a predictor's ability to identify actual dropouts rather than graduates.***

Indeed, a number of studies have tended to examine the relationship of only a single predictor with the odds of high school dropout. For instance, Eide and Showalter (2001) and Stearns, Moller, Blau, and Potochnick (2007) focused on the relationship between grade retention and dropout.

[1] The authors owe a debt of gratitude to Rice undergraduate students Mark Trainer and Chavonte Wright for their helpful assistance compiling an annotated bibliography of much of the dropout literature. Due to the volume of that literature, and due to the fact that high school non-completion is a perennial concern of policymakers, their help was invaluable and went a long way in making the write-up of the background and formation of research questions almost effortless. We also thank the members of the Houston Education Research Consortium's team of junior and senior researchers, as well as conferees at a couple professional annual meetings, for their probing insights of, and suggestions for improvement to, the full draft upon which this brief is based.

Hirschfield (2009) looked at how juvenile arrest was associated with failure to complete high school. Reynolds, Temple, Robertson, and Mann (2001) assessed the long-term consequences of early childhood intervention on high school dropout.

In their review of 44 studies, though, Hammond, Linton, Smink, and Drew (2007) concluded that dropping out of school is related to a variety of factors that can be categorized in four areas, or domains (individual, family, school, and community). They found that accuracy increases when multiple risk factors are considered simultaneously. The most significant factors that are predictive of school dropout among the studies they considered were:

- Over age for grade/retention
- Low achievement
- Poor attendance/high absenteeism
- Low SES

Furthermore, many demographic and other factors (e.g., low bonding to school, being sexually active, and low parental expectations) hypothesized to be predictive of early school leaving are almost entirely mediated by poor academic performance at age 14 (Battin-Pearson et al., 2000). And both deviance and academic competence fully mediated the association of gender and race/ethnicity with high school dropout and both also partially accounted for the family SES-high school dropout relationship (Newcomb et al., 2002). Allensworth and Easton (2005) and Balfanz and Nield (2006) found that GPA and course failures appear to be the “strongest” predictors of dropout, with demographic factors such as gender, race/ethnicity, and family SES offering little additional predictive power.

To move the field toward a more consistent reporting structure, Bowers et al. (2013) advocate calculating a variable’s ***true-positive proportion*** and its ***false-positive proportion***. A more accurate indicator will have a larger true-positive proportion, while a poor or less accurate indicator will have a larger false-positive proportion.

Beyond accuracy, we ought to care about the ***relative strength*** of many covariates in a model. The most efficient allocation of scarce resources will necessarily require being able to pinpoint those factors that are more strongly related to student dropout, all other factors being equal. An early warning system for high school dropout should include variables that are both accurate when considered alone and strongly predictive of dropout when considered in tandem with other accurate indicators.

Assessing dropout predictors in the Houston Independent School District is important given the specific demographics of the district. Expressly, while the findings of previous studies may yield support for targeting particular indicators and establishing early warning systems, the demographics of Houston, with its largely Hispanic and black and economically disadvantaged population, might evince the need for focus on other indicators not previously believed to be predictive of dropping out of high school.

RESEARCH QUESTIONS

We first aim to assess the accuracy of a number of high school dropout predictors. However, we note that while a predictor may be accurate in isolation, once we account for other factors, it may not be strongly related to the outcome under consideration: high school dropout. We therefore also aim to assess the relative strength of all our chosen predictors on the odds of high school dropout. Our formal research questions are:

- Which student- and school-level factors are more accurate than not at predicting high school dropout?
- Accounting for all indicators in a single model, what is the relative strength of the relationship of each predictor to the odds of dropping out?

DATA AND METHODS

Data

Data for this study were provided by the Houston Independent School District (HISD). The nation’s seventh largest school district, HISD collects administrative data on more than 200,000 pre-K through 12th grade students annually. Our sample includes the ninth-grade cohort for the 2009–2010 academic year. The Texas Education Agency (TEA) determined, based on data from the Public Education Information Management System (PEIMS), that 13,687 students were at one time part of HISD’s 2009–2010 ninth-grade cohort. Excluding students who left HISD for reasons other than graduating, receiving their GED, dropping out, being confined to a juvenile detention center, or for whom data errors made it impossible to determine their final status, HISD's final 2009-2010 ninth-grade cohort size for analytic purposes was 11,524. For state accountability reasons, an additional 462 students were excluded under Texas Education Code (TEC) subsection 39.053(g–1), for a second final analytic sample of 11,062. We further limit our analytic sample to include only those students for whom scores were available on the eighth grade Texas Assessment of Knowledge and Skills (TAKS) math and reading assessment. Our final analytic

Table 1. Life table for high school dropout.

Grade	Time Interval	R	D	W	$\hat{p}$	$\hat{S}$
8	[0, 1)	9,009	-	-		1.000
9	[1, 2)	9,009	94	229	0.010	0.990
10	[2, 3)	8,686	162	315	0.019	0.971
11	[3, 4)	8,209	275	552	0.033	0.939
12	[4, 5)	7,382	235	7147	0.032	0.909

Note: R represents the number with the potential of dropping out in an interval, D represents the number of students who dropped out at the end of an interval, W represents the number of students who did not return in the following interval, *p*-hat is the within-interval dropout rate, and *S*-hat is the cumulative survival rate.

sample, then, consisted of 9,009 HISD continuing ninth-grade students nested in 52 high schools.

Measures

Our dependent variable is whether or not a student dropped out anytime during the expected four years of high school. At the beginning of ninth grade, all 9,009 students in our analytic sample have the potential of dropping out within the four years. Of that 9,009 students about one percent, or 94 students, did not return to school after the end of 9th grade. Another 229 students left the district for a reason other than dropout, including graduation, reducing the total number of students exposed to risk of dropping out in 10th grade to 8,686. **Table 1** presents a life table for high school dropout across grades 9 through 12. *R* represents the number of students with the potential of dropping out at the beginning of the interval, or grade, *D* represents the number of students who dropped out at the end of the interval, and *W* represents the number of students who did not return to school in the following interval for a reason other than dropout. The final two columns, $\hat{p}$ and $\hat{S}$, denote the within-interval dropout rate and the cumulative survival rate, respectively.

Our key predictor variables of interest include demographic factors known to be associated with high school non-completion, as well as a number of factors that are amenable to policies aimed at altering them. Among the demographic factors are (1) male gender (a dummy variable for which female gender serves as the reference category); race and ethnicity dummies for (2) black, (3) Hispanic, and (4) Asian or other race students (white is the reference category); disadvantaged status dummies for (5) free or reduced lunch and (6) in poverty (not disadvantaged is the reference category).[2] Among the more malleable factors from a policy standpoint are time-invariant dummies for whether a student (7) was classified as special education at the beginning of ninth grade, (8) was classified as limited English proficient (LEP) at the beginning of ninth grade, (9) was 16 years of age or older at the beginning of ninth grade, (10) had a disciplinary incident in eighth grade, (11) experienced disciplinary action in eighth grade of greater than 10 days, (12) received an F for a course during a grading period in eighth grade, or failed to meet the eighth-grade (13) math standard or (14) reading standard on the Texas Assessment of Knowledge and Skills (TAKS). To account for differences between schools that might affect a student's odds of dropping out, we also include among our predictors variables denoting whether the proportion of (15) black, (16) poor, and (17) at–risk students in a given student's school exceeds the district mean proportion among all high schools. **Table 2** shows, by dropout status and for the total sample, summary statistics for all variables in this study.

Analytic Strategy

Analysis proceeded in three steps. First, we assessed whether and to what extent the variables in this study accurately predicted dropping out. In a review of the literature, Bowers et al. (2013) highlight a major concern related to the inaccuracy of predictors believed to be associated with high school dropout. In the case that some students are misidentified as likely to drop out when in fact they do not drop out, and that other students are not identified as likely to dropout when in fact they do drop out, districts will misallocate limited resources by targeting dropout prevention/intervention

[2] Economic disadvantaged code reporting distinguishes between students eligible to receive reduced or free lunch under the National School Lunch Program (NSLP), determined either through a NSLP application or direct certification from the Texas Department of Agriculture, and students deemed to be economically disadvantaged through other means, such as being recipients of Temporary Assistance for Needy Families (TANF) or Supplemental Nutrition Assistance Program (SNAP); also among this group of students are migrant, runaway, and homeless students. Controlling for other factors, the latter group often evidences greater negative outcomes than the former group. As a consequence, it makes sense to treat each category separately and to properly refer to the latter category as poverty status.

Table 2. Summary statistics by dropout status and for total.	Did Not Drop Out		Dropped Out		Total	
	Mean	SD	Mean	SD	Mean	SD
Male	0.48	0.50	0.57	0.50	0.49	0.50
Race/ethnicity						
White	0.09	0.28	0.03	0.16	0.08	0.28
Black	0.26	0.44	0.39	0.49	0.27	0.44
Hispanic	0.62	0.49	0.57	0.50	0.61	0.49
Asian/other race	0.04	0.19	0.01	0.11	0.04	0.18
Disadvantage status						
Not disadvantaged	0.26	0.44	0.16	0.37	0.25	0.44
Reduced/free lunch eligible	0.56	0.50	0.50	0.50	0.56	0.50
In poverty	0.18	0.38	0.34	0.47	0.19	0.39
Special education	0.10	0.29	0.19	0.39	0.10	0.30
Limited English Proficient (LEP)	0.22	0.60	0.24	0.57	0.22	0.60
Age at start of 9th grade	14.83	0.63	15.50	0.81	14.88	0.67
Had a disciplinary incident in 8th grade	0.28	0.45	0.62	0.49	0.31	0.46
Length of disciplinary punishment in 8th grade	2.77	12.27	11.99	27.10	3.55	14.37
Received an F during a grading period in 8th grade	0.64	0.48	0.87	0.33	0.66	0.48
Failed to meet TAKS 8th grade math standard	0.24	0.43	0.50	0.50	0.26	0.44
Failed to meet TAKS 8th grade reading standard	0.11	0.31	0.25	0.43	0.12	0.33
School % black above district mean	0.32	0.47	0.49	0.50	0.33	0.47
School % in poverty above district mean	0.16	0.36	0.36	0.48	0.17	0.38
School % at-risk above district mean	0.60	0.49	0.81	0.39	0.62	0.49

programs at students who do not require it and denying it to the students who may need it most.

To address this concern, Bowers et al. (2013) suggest a method for comparing the accuracy of dropout indicators. Their method involves calculating a binary predictor's sensitivity (i.e., the ratio of those predicted to drop out of high school to the total who did drop out, or the true-positive proportion) and its specificity (i.e., the ratio of those predicted to drop out of high school who graduated, or the false-positive proportion) and plotting the former against the latter. Known as the relative operating characteristic (ROC) plot, the utility of this procedure is that it provides a visual means of assessing the accuracy of a dropout indicator. If one imagines a forty-five degree line drawn from (0, 0) on the Cartesian plane, any indicator falling perfectly on that line would indicate equal values of true-positives to false-positives. Less accurate indicators will maximize the false-positive proportion relative to the true-positive proportion and will fall below the forty-five degree line, and more accurate indicators will maximize the true-positive proportion relative to the false-positive proportion and will fall above the forty-five degree line. **Figure 1** shows a contingency table in which the columns denote the event experienced by a student (e.g., dropped out or graduated) and the rows denote the dropout indicator and whether it predicted dropout or graduation.

Second, we estimated a set of discrete-time logit hazard models. Models were parameterized without an intercept so that the estimates for the four time periods during high school (i.e., ninth through 12th grade) referred to the odds of dropping out of high school in the specific period. The covariate effect on the odds was assumed to be the same in all periods, or proportional, so the estimates of all covariates, when exponentiated, can be interpreted as the multiplicative effects of the odds in any time period.

Finally, following methods described by Menard (2011), we calculated standardized regression coefficients so as to be able to assess the relative importance of each predictor used in this study. We rank each of the covariates in order of the strength of its relationship to the dependent variable. Rank ordering is based on the absolute value of the standardized coefficient.

Steps two and three are explained in greater detail in **Appendix A**.

		Event: Dropout	Event: Graduate	
Predictor	Dropout	a True-Positive Correct	b False-positive Type I-Error	*a + b*
	Graduate	c False-negative Type II-Error	d True-negative Correct	*c + d*
		a + c	*b + d*	*a + b + c + d = N*

Figure 1. Event table for calculating dropout contingency proportions.

RESULTS

Table 3 shows values for sensitivity and specificity for each dropout indicator used in this study. When we look at the percentage of all dropouts who had the dropout indicator, shown in the third column under sensitivity, seven of the covariates have values at or exceeding 50 percent. Among these factors are male, Hispanic, free or reduced lunch, had disciplinary issues in eighth grade, failed period/course in eighth grade, eighth grade TAKS math standard not met, and school percent at–risk greater than district mean. Since more accurate predictors will maximize the percentage of dropouts who had the dropout flag, orthe value for sensitivity, while minimizing the percentage of graduates who had the dropout flag, or the value for specificity, the corresponding values under column 5 should be smaller than those in column 4. This is true for five of the seven factors just mentioned. Both Hispanic and free or reduced lunch are poor predictors of high school dropout in HISD. Status as Asian or other race student was also a poor predictor of dropout, bringing the total number of covariates in this study for which the value on specificity exceeds the value on sensitivity to three. The negative values for these covariates, shown in the final column of Table 3, demonstrate this very clearly. All other covariates in this study are better predictors than not of high school non-completion.

Figure 2 shows the relative operating characteristic (ROC) of the dropout indicators, plotted as the sensitivity against the specificity

Table 3. Values for precision, sensitivity, specificity, and the difference between sensitivity and specificity.

ID	Dropout Indicator	Sensitivity *a* /(*a* + *c*)	1-Specificity *b* /(*b* + *d*)	Sensitivity - 1-Specificity
1.	Male	0.57	0.48	0.09
2.	Black	0.39	0.26	0.13
3.	Hispanic	0.57	0.62	-0.05
4.	Asian/other race	0.01	0.04	-0.03
5.	Free/reduced lunch	0.50	0.56	-0.06
6.	In poverty	0.34	0.17	0.17
7.	Special Education	0.19	0.10	0.09
8.	Limited English Proficient (LEP)	0.14	0.09	0.05
9.	Age at start of 9th grade 16 or older	0.27	0.05	0.22
10.	Had a disciplinary incident in 8th grade	0.62	0.28	0.34
11.	Disciplinary punishment greater than 10 days in 8th grade	0.19	0.04	0.15
12.	Failed Period/Course in 8th grade	0.86	0.36	0.50
13.	8th TAKS math standard not met	0.50	0.24	0.26
14.	8th TAKS reading standard not met	0.25	0.11	0.14
15.	School % black greater than district mean	0.49	0.32	0.17
16.	School % in poverty greater than district mean	0.36	0.16	0.20
17.	School % at-risk greater than district mean	0.81	0.60	0.21

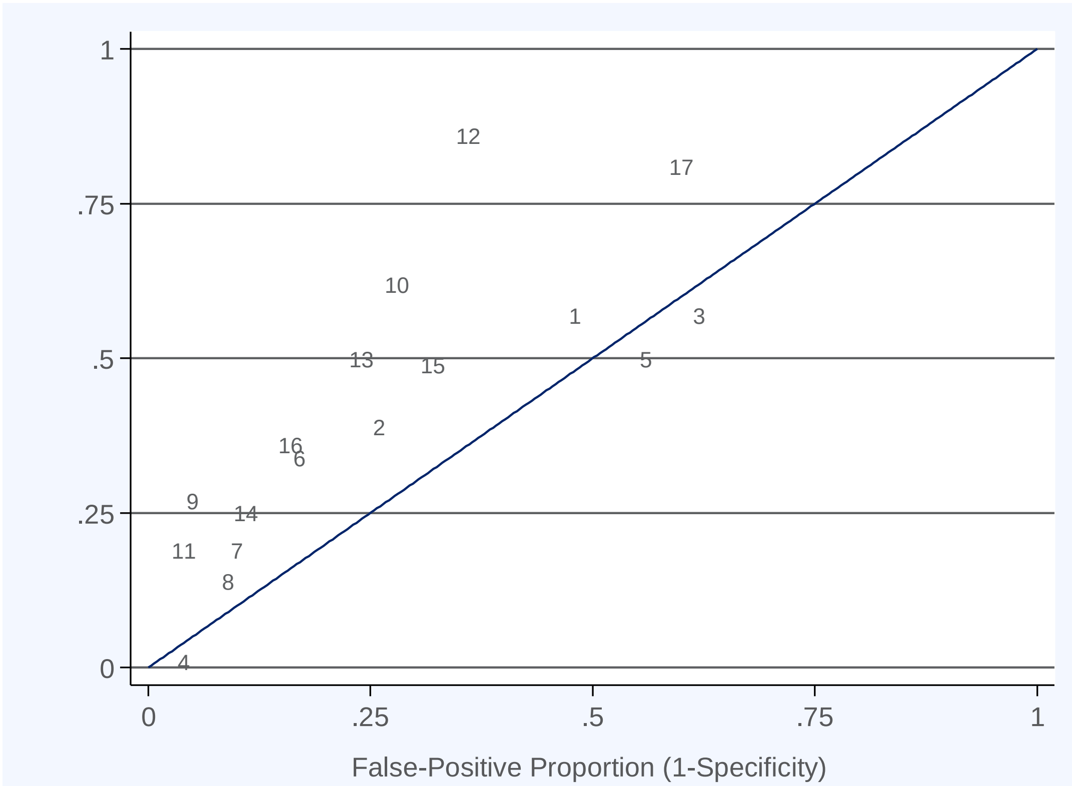

Figure 2. Relative operating characteristics (ROC) of dropout indicators, plotted as the true-positive proportion against the false-positive proportion (see Table 3 for indicator description).

proportion. As was just stated with regard to Table 3, only three items are poorer predictors than not of high school dropout (i.e., they fell below the forty-five degree line) while all the remaining covariates are better predictors than not of high school dropout (i.e., they fell above the forty-five degree line). Variable IDs 10 and 12, which correspond to the IDs in Table 3, appear to be the closest to (0,1) and are the most accurate predictors of high school dropout using this method of predictor assessment.

To understand how each covariant relates to the dependent variable, net of other factors, we turn to our discrete-time logit models, shown in **Table B1** (page 11). Model 1, the null model, presents the unadjusted, or marginal, hazards of dropout in each time period or grade. Similar to the yearly dropout rates shown in the life table (Table 1), we see that about 1 percent of those with the potential of dropping out in grade 9 did so. This baseline increases to approximately 2 percent of the remaining set of students with the potential of dropping out in grade 10, and to about 4 percent in both grades 11 and 12. There appears to be some modest between-school variability in the dropout propensity as given by $\widehat{\sigma_u^2}$, which is statistically significant at the .05 level.

Model 2 adjusts the annual baseline hazards with the addition of the demographic factors for gender, race/ethnicity, and disadvantaged status, and the covariates for school-ascribed status as either special education or LEP. The odds of dropping out in any given grade ranges from about .5 percent to 1.5 percent, but is 31 percent higher among males relative to females, holding other covariate effects constant. Relative to whites, blacks and Hispanics have statistically significant higher odds of dropping out in any given year of high school, by 76 percent and 81 percent, respectively. Asians and other race students in HISD are slightly less likely to drop out of school than whites, net of other factors, but their lower odds are not different from zero. Students who qualified for free or reduced lunch are as likely to remain in school as are their peers from advantaged families. Students from poor families, on the other hand, have 83 percent higher odds of dropping out in any year of high school than students from advantaged families. Finally, while status as special education is not associated with higher odds of dropping out, all other things being equal, status as LEP is; current LEP students are 55 percent more likely than their non-LEP peers to drop out in any given year of high school. Again, $\widehat{\sigma_u^2}$ suggests a great deal of between-school variability in the odds of dropping out.

Moving on to Model 3, which adds the remainder of the controls, the greater odds of high school non-completion associated with male gender, black and Hispanic race, and LEP status are completely attenuated. Only the coefficient for poverty status remained statistically significant; relative to advantaged students, poor students still have 80 percent higher odds of dropping out in any given year of high school, net of the additional factors controlled for under Model 3. Higher odds of dropout are associated with the majority of the other covariates, including over age for grade at the start of ninth grade (336 percent), whether the student had a disciplinary incident in eighth grade (124 percent), whether the length of disciplinary action in eighth grade exceeded 10 days (1 percent), whether the student received an F during a grading period in eighth grade (116 percent), and whether the student failed to meet the TAKS eighth grade math (65 percent) or reading (36 percent) standard. Two of the three school-level measures are associated with higher odds of dropping out. Students who attended a school in their freshman year in which the percent in poverty was at or exceeded the district high school mean have 84 percent higher odds of dropout than their peers at schools where the percent in poverty is below the district mean. Similarly, students whose freshman year school's percentage at-risk was at or above the district mean are 43 percent more likely to drop out in any given year of high school than those whose freshman year school's percentage at-risk was below the district mean. Interestingly, the extra controls included in Model 3 revealed no significant amount of between-school variability in the odds of dropout as shown by $\widehat{\sigma_u^2}$.

Table 4 compares the unstandardized (non-exponentiated) and fully standardized coefficients for each dropout indicator. As the latter constitutes a means for assessing the relative strength of each indicator, we also rank the covariates to demonstrate their relationship to the odds of dropping out. In this study, over age for grade had the strongest relationship with student dropout, net of all the other covariates, followed by the receipt of an F during any grading period in the eighth grade, whether the student had a disciplinary incident in eighth grade, and so on.

DISCUSSION AND RECOMMENDATIONS

The aim of this study was to assess, using student- and school-level data from HISD, both the accuracy and strength of a number of predictors believed to be related to high school dropout. Findings showed that most of the indicators under consideration were more accurate than not at predicting the odds of failing to complete

Table 4. Relative rank of covariates.

Variable	Unstandardized Coefficient	Standardized Coefficient	Rank
Age at start of 9th grade 16 or older	1.485	0.190	1
Received F during grading period in 8th grade	0.771	0.070	2
Had a disciplinary incident in 8th grade	0.770	0.068	3
Failed to meet TAKS 8th grade math standard	0.571	0.048	4
School % at-risk above district mean	0.455	0.043	5
In poverty	0.564	0.042	6
School % in poverty above district mean	0.558	0.041	7
School % black above district mean	0.232	0.021	8
Failed to meet TAKS 8th grade reading standard	0.310	0.020	9
Length of disciplinary punishment in 8th grade > 10 days	0.007	0.019	10
Hispanic	0.166	0.016	11
Reduced/free lunch eligible	0.125	0.012	12
Male	0.106	0.010	13
Asian/other race	-0.111	-0.004	14
Special education	-0.059	-0.003	15
Black	-0.006	-0.001	16
Limited English Proficient (LEP)	-0.009	-0.001	17

high school. That is, for all but three of the predictors used in this study (Hispanic, Asian or other race, and reduced- or free-lunch eligible), there were more dropouts who were flagged to drop out than there were graduates who were flagged to drop out. When all variables were entered into a logistic regression analysis, most were statistically significantly related to the odds of dropping out. Net of covariates, there was no statistically significant relationship between the odds of dropout and male gender, race/ethnicity, special education status, limited English proficient status, or whether the school's proportion of black students exceeded the district mean.

Based on the standardized logistic regression the relative rank of covariates revealed that being overage at the start of ninth grade had the strongest relationship with the propensity to drop out of high school, followed by receipt of an F during any period in eighth grade, recording of a disciplinary incident in eighth grade, and failure to meet the TAKS eighth-grade math standard. Again, race, gender, and reduced- or free-lunch eligible ranked low with regard to their relationship to the odds of dropping out during high school.

We note a few limitations. First, we were unable to control for a number of other factors also associated with increased odds of early school leaving, including student motivation, family involvement and mobility, the instructional quality of the school, peer environment and influence (Cairns, Cairns, & Neckerman, 1989; Ensminger, Lamkin, & Jacobson, 1996), just to name a few. The Theory of Planned Behavior (TPB) is relevant to the topic of student motivation since intentions are believed to predict behavior via attitudes, subjective norms, and perceived control. Davis, Ajzen, Saunders, and Williams (2002) provide the only study that uses TPB to predict dropout.

Second, the predictors in this study merely indicate the odds of dropping out and are not the cause, per se, of dropping out. The causes of high school non-completion are varied and more than likely exert an influence earlier in the life course than the period between eighth and ninth grades. As our data include administrative student-level and school-level data not supplemented by instruments designed to assess student psychology, we are limited in saying our predictors are driving dropout since psychological factors could be the primary driving force behind students' greater likelihood of dropping out.

Finally, this study focuses on a single cohort and the findings here may be specific to that cohort alone. That said, we did run analyses on the 2011–2012 graduating class (i.e., the 2008–2009 ninth grade cohort) as well and the findings were essentially the same as those shown here, but two years does not necessarily a trend make; analyses for future cohorts may reveal different results than the results shown in this study.

Despite these limitations, the findings of this study are consistent with the extant literature on early school leaving (see Hammond et al.'s [2007] review). Given this fact, and given that the most consequential factors associated with high school non-completion are student-level school-malleable factors, suggestions for intervention to prevent the likelihood of dropout can be offered with a high level of confidence. Specifically, to forestall high school non-completion within HISD, intervention should at the very least be targeted at bringing students up to grade level. Identifying students who are over age at the start of high school and working with them to overcome deficiencies in the subject areas that may have caused them to be retained in grade in the first place should be a top goal of the district. Because high school-age students are so close to adulthood, however, those over age for grade are going to be that much closer. It may therefore be very challenging for schools and teachers to get those students to fully re-engage with their academic studies. An optimal intervention would be for the district to strive to keep students at grade level throughout the elementary and secondary years so that there are fewer students over age for grade at the start of ninth grade. This presents its own unique challenges of course. As amenable as academic difficulty is to school- and teacher-level efforts, home and community factors are often not easily overcome.

Aiming to bring students up to, or keep them at, grade level would, by extension, target the second and fourth strongest dropout indicators shown in this study (i.e., receipt of an F during any grading period in eighth grade and failed to meet the TAKS eighth grade math standard), though, again, outside factors may be difficult to overcome. Notwithstanding these difficulties, giving students the opportunity to remedy past or present shortcomings nonetheless remains an achievable aim. Making summer school a requirement for elementary and middle school students who have failed a course, or who might, under a different regime, be retained in grade, may go a long way in preventing dropout in high school. This is somewhat conjectural, but the intuition behind it lies in the fact that the effective use of instructional time is of utmost importance in helping struggling students keep up with their peers. Since retention often duplicates a whole year of schooling, it is obvious that this may not be the most effective use of instructional time.

Finally, regarding the third strongest indicator of high school dropout, whether a student had a disciplinary incident in eighth grade, it is recommended that schools have staff on hand equipped to assist students with managing their behavioral dispositions. These staff need not be limited to teachers, but can include counselors and therapists with special skills to deal with the more extreme behavior situations. For students who commit offenses so extreme that they are assigned to an alternative education program within the district, greater effort should be made to reintegrate said students back into their regular school population once they exit the alternative education program. Because such students often misbehave as a consequence of factors far outside of school control and related to their lives at home and in their neighborhoods, treatment of students by credentialed professionals such as school or clinical psychologists may be needed on an ongoing basis to prevent repeated incidences.

REFERENCES

Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago: Consortium on Chicago School Research, University of Chicago.

Anderson, D. M. (2014). In school and out of trouble? The minimum dropout age and juvenile crime. *The Review of Economics and Statistics, 96*(2), 318–331.

Arum, R., & Lafree, G. (2008). Educational attainment, teacher-student ratios, and the risk of adult incarceration among U.S. birth cohorts since 1910. *Sociology of Education, 81*(4), 397–421.

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

Bowers, A. J., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal, 96*(2), 77–100. doi: http://dx.doi.org/10.1353/hsj.2013.0000

Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early school dropout: Configurations and determinants. *Child Development*, *60*, 1437-1452.

Davis, L. E., Ajzen, I., Saunders, J., & Williams, T. (2002). The decision of African American students to complete high school: An application of the Theory of Planned Behavior. *Journal of Educational Psychology, 94*, 810–819.

Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review, 20*(6), 563–576.

Ensminger, M. E., Lamkin, R. P., & Jacobson, N. (1996). School leaving: A longitudinal perspective including neighborhood effects. *Child Development, 67*, 2400-2416.

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.

Hammond, C., Linton, D., Smink, J., & Drew, S. (2007). Dropout risk factors and exemplary programs: A technical report. Clemson, South Carolina: National Dropout Prevention Center.

Hirschfield, P. (2009). Another way out: The impact of juvenile arrests on high school dropout. *Sociology of Education, 82*(4), 368–393.

Menard, S. (2011). Standards for standardized logistic regression coefficients. *Social Forces, 89*(4), 1409–1428.

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

Newcomb, M. D., Abbott, R. D., Catalano, R. F., Hawkins, J. D., Battin-Pearson, S., & Hill, K. (2002). Mediational and deviance theories of late high school failure: Process roles of structural strains, academic competence, and general versus specific problem behaviors. *Journal of Counseling Psychology, 49*(2), 172–186. doi: 10.1037/0022-0167.49.2.172

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association, 285*(18), 2339–2446.

Rumberger, R., & Lim, S. A. (2008). Why Students Drop Out of School: A Review of 25 Years of Research *California Dropout Research Project*. Santa Barbara, California: University of California Santa Barbara.

Rumberger, R. W., & Lamb, S. P. (2003). The Early Employment and Further Education Experiences of High School Dropouts: A Comparative Study of the United States and Australia. *Economics of Education Review, 22*(4), 353–366.

Stearns, E., Moller, S., Blau, J., & Potochnick, S. (2007). Staying back and dropping out: The relationship between grade retention and school dropout. *Sociology of Education, 80*(3), 210–240.

This is an abbreviated version of a much longer research study written for peer review. For additional information on the findings presented here, or to obtain the full peer-review version of this research brief, contact the Houston Education Research Consortium at 713-348-2802 or email herc@rice.edu.

Appendix A

This appendix provides information on the analytic strategy used in this study.

Discrete-time Logit Hazard Modeling

We estimated three discrete–time hazard models for the dropout data, parameterizing our models without an intercept so that the estimates refer to the odds of dropping out in each time interval. In the general model we defined a set of dummy variables, D_{i1}, …, D_{i4}, corresponding to the time intervals, or grades, and specified the following model:

$$logit(p_{ij}) = \alpha_1 D_{i1} + \alpha_2 D_{i2} + \alpha_3 D_{i3} + \alpha_4 D_{i4} + \mathbf{x}'_{ij}\beta + u_i \quad (1)$$

In this equation p_{ij} denotes the conditional probability that a student *i* drops out during interval *j* having been at–risk of drop out after the beginning of interval j. Estimates for the dummies, $D_{i1}, \dots, D_{i4}$, represent the baseline hazard dropout for interval j. We included **x** to account for observed sources of variation in the conditional odds as measured by the predictors in this study. The level–two residual, u_i, is assumed to be normally distributed with mean 0 and variance σ_u^2, and independent of **x**.

An assumption of discrete-time hazard models is proportionality in the odds across time. If a covariant effect β adjusts the logit of the hazard in the same way in all time periods, then exp(β) can be interpreted as the multiplicative effects of the odds of dropout in any interval *j*. We tested the proportional odds assumption by multiplying the covariates with the period dummies and found no evidence of time-varying effects.

Model 1, the null model, included only estimates for the time period dummies. The estimates, therefore, show the marginal hazards of dropping out in each grade. Model 2 added demographic factors for gender, race/ethnicity, and disadvantaged status, as well as for school-ascribed statuses in special education and LEP courses. Finally, Model 3 included the remaining dropout indicators outlined in the previous section.

Calculating the Fully Standardized Logistic Regression Coefficient

While many make the mistake of comparing covariate effects exp(β) as though they conveyed information about the magnitude of variables' effects, the fact is that they impart the same information as the unstandardized logit coefficients. Others have attempted to compare the relative strength of predictors based on the level of statistical significance, but statistical significance, being dependent on sample size, reveals nothing about a predictor's substantive significance. Menard (2011) provides standards for calculating fully standardized logit coefficients to allow for the proper comparison of the relative strength of two or more model covariates. Standardized logit coefficients are useful for yielding meaningful interpretations to variables with no natural metric, such as many of those used in this study. They are also the only appropriate means of comparing variables measured in different metrics. We therefore estimated fully standardized logistic regression coefficients, which we quantified as

$$b_M^* = b(s_x)R/s_{logit(\hat{p})}, \quad (2)$$

where *b* is the unstandardized logistic regression coefficient, s_x is the standard deviation of the independent variable x, and *R* is the correlation between the dependent variable p and the predicted values of $\hat{p}$.

Using the fully standardized logistic regression coefficients, we ranked each of the variables in order of the strength of its relationship to the dependent variable. Rank ordering was based on the absolute value of the standardized coefficient.

Appendix B

Table B1. Estimates from discrete-time logit models of high school dropout.

	Model 1	Model 2	Model 3
		Baseline Hazard	
	$\widehat{e^{\alpha}}$	$\widehat{e^{\alpha}}$	$\widehat{e^{\alpha}}$
[0, 2)	0.012***	0.004***	0.001***
[2, 3)	0.022***	0.008***	0.002***
[3, 4)	0.042***	0.016***	0.004***
[4, 5)	0.043***	0.017***	0.005***
		Multiplicative effects	
		$\widehat{e^{\beta}}$	$\widehat{e^{\beta}}$
Male		1.306***	1.091
Race/Ethnicity (ref. = White)			
Black		1.763*	0.936
Hispanic		1.812*	1.088
Asian/other race		0.931	0.938
Disadvantaged status (ref. = Not disadvantaged)			
Reduced/free lunch eligible		1.055	1.15
In poverty		1.834***	1.803***
Special education		1.076	0.932
Limited English Proficient (LEP)		1.548***	1.036
Age at start of 9th grade 16 or older			4.355***
Had a disciplinary incident in 8th grade			2.243***
Length of disciplinary punishment in 8th grade exceeded 10 days			1.011***
Received an F during a grading period in 8th grade			2.162***
Failed to meet TAKS 8th grade math standard			1.650***
Failed to meet TAKS 8th grade reading standard			1.364**
School % black above district mean			1.168
School % in poverty above district mean			1.841**
School % at-risk above district mean			1.432*
$\widehat{\sigma_u^2}$	2.053*	1.559*	0.088
log *L*	-3397.01	-3351.28	-3004.38
n (person years)	33300	33300	33300
df	4	12	21

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; two-tailed tests.