An Analysis of Dropout Predictors within a State High School Graduation Panel

Bobby J. Franklin, PhD
Assistant Professor of Research
School of Education
Department of Teacher Education & Leadership
Mississippi College
Clinton, MS

Stephen B. Trouard, PhD
Assistant Professor of Management
School of Business
Mississippi College
Clinton, MS

Abstract

A state high school graduation panel ($n = 31,641$) was used to examine four eighth grade predictors of dropouts. The dependent variable was completion status coded as graduate or dropout. The independent variables were age, attendance, gender, and test score. The data were analyzed using logistical regression. Results indicate that the four variables are significant predictors with age and gender being the best. The issue of age is important for policy decisions regarding promotion.

Dropout rates are one of the standards by which school systems are deemed successful. Many schools have such dismal records of success that they carry the label "dropout factories" (Balfanz & Legters, 2004). The impetus to improve the dropout condition carries with it the need for better methods that identify and monitor potential dropouts, such as longitudinal databases and dropout warning systems (Means, Padilla, & Gallagher, 2010).

The establishment and maturity of state longitudinal data systems affords educators and researchers new opportunities and ways to analyze education data, especially in the area of high school completion. Conditions or characteristics that can describe the success of students during their high school years can be scrutinized to a greater degree. In conjunction with longitudinal data systems, many advocacy groups have called for dropout warning systems (Pinkus, 2008) as a tool to help hold back the dropout tide.

Purpose of the Study

The purpose of this study is to expand on the longitudinal perspective of early dropout identification through the use of a state longitudianal database rather than national survey data or district data. Logistical regression was used to determine the effectiveness of four independent variables (age, attendance, gender, and test score) for predicting high school dropouts.

Dropout Indicators

Identification requires knowledge of variables that make predicting dropouts possible. Several reviews of the literature exist that have examined the linkage between prior conditions and dropping out of school (Hunt, 2008; Hupfeld, n.d.; Jerald, 2007; Rumberger & Lim, 2008). Jerald (2007) and Hupfeld (n.d.) categorized variables into groups such as demographic background, family factors, adult responsibilities and education experiences. Rumberger and Lim (2008) categorized dropout factors as either individual or institutional based on an examination of 203 published works. These two categories included academic achievement, attendance, mobility, age, engagement, delinquency, peers, work attitudes, self-perceptions, prior training and demographics, with academic achievement (test scores and grades) topping the list of variables studied. Nearly 200 studies included gender as part of their analyses, but the results often depended on other variables present in the statistical models (Rumberger & Lim).

Timeframes

The ninth grade is viewed as a focal year to begin the monitoring process (Allensworth & Easton, 2007). This is the time students begin to earn credits for graduation and the time more students experience failure (Kennelly & Monrad, 2007). There are instances where indicators from elementary and middle school grades have been used to predict high school dropouts (Neild, Balfanz, & Herzog, 2007; Logan, 2010; Rumberger & Lim, 2008). Studies with elementary and middle schools address the impact non-promotion and school transitions have on the dropout process. These studies also indicate that dropout factors manifest themselves long before high school (Logan, 2010). While monitoring student progress in high school is critical, not scrutinizing student progress during the elementary and middle school years is irresponsible.

Data sources

Longitudinal studies have primarily used data from federal datasets, selected schools, or large school districts. Rumberger and Lim (2008) found that between 1982 and 2008 about 79% of the analyses examined used some type of national longitudinal dataset. Analyses using state data have tended to focus on certain subpopulations (Gregg, 2010; Logan, 2010; Rumberger & Lim, 2008). Other studies have centered on large urban districts such as Philadelphia (Neild et al., 2007), Chicago (Allensworth & Easton, 2007) and Baltimore (Entwisle, 2005). No studies

were found that examined dropout predictors using a state-level panel developed according to national guidelines.

Summary

There is a need to examine the high school dropout issue using a compreshesive data system containing reliable, unbiased data (Logan, 2010). While national longitudinal surveys provide valuable information, verification of previous findings using a data system designed to track dropouts is essential to the contuned understanding of the dropout phenomon. Schools will ultimately benefit from knowing what combinations of 8th grade data generate stronger relationships and higher predictive value (Ormeseth, 2010) and will avoid wasting valuable resources (Bowers, 2010). This article addresses the question, "Are the variables age, attendance, gender, and test score collected using a state longitudinal database in the eighth grade for a ninth grade cohort effective in predicting dropouts during the high school years?"

Method

A state data system was used to create a panel file to identify graduates and dropouts. Students who entered high school as ninth graders were tracked across the ensuing four years of high school to determine their exit status. Students within the panel were placed into two groups: (a) graduates, students who received a diploma and (b) dropouts, students who met the NCES definition (Stillwell & Sable, 2013) of dropout.

Sample

The panel was populated with first time 9th grade students for a corresponding four-year high school period. Legitimate transfers to other educational facilities were removed. Students who graduated from a public school with a high school diploma and students who dropped out of the public school system without enrolling in another education institution composed the sample used in this study (n= 31,641).

Variable Definitions

Age. Four independent variables were examined for their relationship to the dependent variable completion status (graduates vs. dropouts). The independent variables are age, attendance, gender, and test score for the year prior (eighth grade) to the initiation of the cohort. Age was calculated to the nearest tenth of a year as of one year prior to the first day of the panel. For the purposes of this calculation, the date of August 14 of the eighth grade year was used to provide some consistency since actual start dates vary within a few days of this date across the state.

Attendance. Attendance is the attendance rate during the year prior to entering the cohort. The attendance rate is the number of days the student was present at school divided by the number of days the student was enrolled converted to a percentage. For students who attended more than one school, multiple counts were summed across all schools in the state to calculate a single rate.

Gender. Gender was the only dichotomized independent variable in the analysis and the only one of the four not affected by time. These data were entered in the database as "M" for male and "F" for female. For this study, the female students were assigned the code "1" and male students were assigned the code "0".

Test score. Test score is also based on data collected during the year preceding the establishment of the cohort. This is the eighth grade exit exam administered by the state. The state assessment for grade eight is a criterion-referenced test that is used as a gate-keeper exam for entry into high school. For this analysis, the English Language Arts (ELA) test score was used as a measure of student performance.

Completion status. A categorical dependent variable (completion status) was created using the exit codes from the database. Completion status was assigned the values of graduate or dropout. Students, who completed high school at the end of the traditional four-year period with a diploma, were identified through the tracking system as graduates. Dropouts were those students that exited high school during the same four-year period without receiving a high school diploma, certificate of completion, or did not transfer to another educational institution.

Statistics

This research utilized the statistical technique logistic regression to measure the impact of age, attendance, gender, and test score upon the dropout rate of high school students in the public school system. Logistic regression is a form of regression analysis that is analogous to multiple regression, but allows for the estimation of a binary categorical dependent variable based upon one or more independent predictor variables (Hair, Black, Babin, & Anderson, 2010). Logistic regression permits the researcher to use non-normal data. It also permits the usage of coded independent variables.

<div align="center">

Results

</div>

Descriptive Statistics

The sample consisted of 31,641 eighth-grade students that entered high school the following year. At the conclusion of the four-year high school cycle, 5,430 (17.2%) were classified as dropouts and 26,211 (82.8%) were identified as graduates. The overall gender composition was 46.9% males and 53.10% females. The subgroup descriptions reveal a greater proportion of males (57.6%) within the dropout group and a higher proportion of females (55.4%) among the graduates.

The average age of eighth-graders at the beginning of the school year was 13.7 with dropouts exhibiting a higher average (14.4) than the graduates (13.6). The age difference between these two groups exhibits a large effect size (Cohen's d = 1.14) based on Cohen's (1988) convention for a large effect (Cohen's $d \geq .80$). The overall eighth-grade ELA test score average was 32.8. The dropout average score was 30.0 compared to the graduate average score of 33.3. This also produces a large effect size (Cohen's d = .94). Dropouts had an eighth-grade attendance rate of 90.2% while the graduates show an attendance rate of 95.6%. Again, the effect size was large (Cohen's d =.83). The descriptive statistics for each independent variable are presented in Table 1.

Logistic Regression

Just like multiple regression, logistic regression utilizes a base model for comparison purposes. Logistic regression uses the likelihood value instead of the sums of squares used in multiple regression. Model estimation fit is measured using the log of the likelihood value multiplied by -2. This value, referred to as -2LL, has a minimum assessment of zero, and represents an ideal model fit. For the comparison of different equations to determine the best model fit, a lower -2LL value represents the best model fit. The -2LL value for the base model in this research was 29010.639. The -2LL value for the hypothesized model including the four independent variables was 19167.647. This 33.90% decrease in the -2LL value displays that the hypothesized model is an improvement to the base model.

Table 1

Descriptive Statistics for Continuous Independent Variables

Variables	*M*	*SEM*	*SD*	Variance
Age	13.699	.0038	.6696	.448
Test Score	32.753	.0199	3.540	12.533
Attendance %	94.667	.0300	5.334	28.453

Note. n = 31,641. Gender is a dichotomous variable and not represented in this table.

The classification table is used to measure how well student completion status is predicted. This measure of predictive accuracy also enables the comparison of the base model to the hypothesized model. The base model without any predictor variables correctly classified 82.8% of the cohort members. With the predictor variables added, the hypothesized model correctly identified 88.0% of the students. The increase in the overall prediction percentage suggests that the hypothesized model is a robust predictor of students' completion status.

The logistic regression calculations determine which of the four independent variables (age, attendance, gender, and test score) were significant predictors of high school dropouts. Significance tests for the independent variables are presented in Table 2. The Wald statistic for each of the variables tests the statistical significance of each of the independent variables. The

results of the tests show that each independent variable is a significant predictor of student completion status.

Table 2

Significance Tests for the Independent Variables

Variables	*B*	*S.E.*	Wald	Exp(B)
Age	1.391	.028	2421.793	4.020
Gender	-.263	.038	47.612	.769
Test Score	-.192	.006	992.618	.825
Attendance	-.146	.004	1505.857	.864
Constant	-.996	.597	2.782	.369

Note: $df = 1$ and the significance level is $p < .001$ for all variables except the constant which is $p < .010$.

Lastly, the odds ratios [Exp (B)] were calculated. The odds-ratio displays the strength of the relationship between a predictor variable and the response variable. The ratio can produce results ranging from 0 to infinity with a result of 1 meaning no relationship between the two variables. For results less than 1, a change in the predictor variable will result in a decline in the response variable. For results greater than 1, a change in the predictor variable will result in an increase in the response variable. The odds ratio of 4.020 for age shows that a student is 302.0% more likely to drop out for every year increase in age when they enter the eighth grade. The odds ratio of 0.769 for gender shows that a student is 23.1% less likely to drop out if they are female. The odds ratio of 0.825 for test scores indicates that a student is 17.5% less likely to drop out for every 10-point increase in their eighth-grade English Language Arts test score. The odds ratio of 0.864 for attendance shows that a student is 13.6% less likely to drop out for every 1% increase in their eighth grade attendance rate. See Table 2 for these results.

Conclusion

Longitudinal data systems can provide valuable information about the progression of students across time along with other aspects of the health of the school system. Administrators should take advantage of these data to identify potential dropouts early in the process. The variables examined here are easily obtained from standard data systems and can feed directly into early warning systems. Dropout early warning systems need not be complicated, but research needs to continue to increase model accuracy.

The results of this study indicate that each independent variable examined was an effective predictor of student completion status. Among the four variables examined, the age of the student upon entrance into the eighth grade was the strongest predictor. This finding has serious implications for student retention practices within public schools. Once a student is retained, especially for academic reasons, the probability that the student will become a dropout increases appreciably. It is not surprising to see gender as a valuable predictor given that males have tended to drop out of school at higher rates than females. Likewise, students with higher test scores in the eighth grade and higher attendance rates are less likely to dropout.

This analysis underscores the usefulness of applying diagnostic methods long before a student enters high school to identify potential dropouts. Waiting until students enter high school to monitor their progression does nothing but delay intervention for those that currently need help. Postponing intercession into the dropout process is not a logical course of action. With the development and implementation of state longitudinal data systems, the monitoring process should begin much earlier to provide more time for correcting the situation.

Additional research is needed to determine if these same results can be obtained with earlier cohorts and if other variables can be identified. Research should also examine those students that do not fall into either of the two completion categories used in this study. Further analysis of state longitudinal data systems will give a better understanding of the disengagement process that leads to dropping out of school and provide opportunities for early intervention. A more beneficial application would be to use the data to identify curricular and instruction flaws to eliminate the need for student remediation and keep students on track to graduate.

References

Allensworth, E., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures and attendance in the freshman year.* Chicago, IL: Consortium on Chicago School Research.

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* The Johns Hopkins University, Baltimore, MD: Center for Research on the Education of Students Placed At Risk.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Leadership*, *103*(3), 191-207.

Entwisle, D. A., Alexander, K. L., & Olson, L. S. (2005). Urban teenagers: Work and dropout. *Youth & Society*, *37*(3), 3-32.

Gregg, W. S. (2010). *Middle school student records as dropout indicators* (Doctoral dissertation). Retrieved from ERIC. (ED520143)

Hair, J. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis.* Upper Saddle River,NJ: Prentice Hall.

Hunt, F. (2008). *Dropping out from school: A cross-country review of literature* (Research Monograph No. 16). Falmer, United Kingdom: University of Sussex, Centre for International Education. Retrieved from http://www.create-rpc.org/pdf_documents/PTA16.pdf

Hupfeld, K. (n.d.). *A review of the literature: resiliency skills and dropout prevention.* Retrieved from Scholar Centric website: http://www.scholarcentric.com/research/ SC_Resiliency_ Dropout%20Prevention_WP_FNL.pdf

Jerald, C. (2007, April 5). *Keeping kids in school: What research tells us about preventing dropouts.* Retrieved from The Center for Public Education website: http://www.centerforpubliceducation.org/site/apps/nlnet/content3.aspx?c=lvIXIiN0JwE&b=5113477&content_id={07DA7811-BB77-46DA-A9D6-8E589B8E2A0F}&notoc=1

Kennelly, L., & Monrad, M. (2007, October). *Approaches to dropout prevention: Heeding early warning signs with appropriate interventions.* Retrieved from Better High Schools website: http://www.betterhighschools.org/docs/NHSC_ApproachestoDropoutPrevention.pdf

Logan, L. E. (2010). *Identifying middle school students at risk for dropping out of high school* (Doctoral dissertation). Liberty University, Lynchburg, VA. Retrieved from http://digitalcommons.liberty. edu/cgi/viewcontent.cgi?article=1414&context=doctoral

Means, B., Padilla, C., & Gallagher, L. (2010). *Use of education data at the local level: From accountability to instructional improvement.* Wahington DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/tech/use-of-education-data/index.html

Neild, R. C., Balfanz, R., & Herzog, L. (2007). An early warning system. *Educational Leadership*, *65*(3), 28-33.

Ormeseth, B. (2010). *Eighth grade indicators of successful transition to high school* (Doctoral dissertation). Cardinal Stritch University, Milwaukee, WI .

Pinkus, L. (2008). *Using early-warning data to improve graduation rates: Closing cracks in the education system.* Washington, D C: Alliance fo Excellent Education.

Rumberger, R., & Lim, S. (2008, October). *Why students drop out of school: A review of 25 years of research* (California Dropout Research Project Report #15). Retrieved from California Dropout Research Project: http://www.cdrp.ucsb.edu/pubs_reports.htm#15

Stillwell, R., & Sable, J. (2013). *Public school graduates and dropouts from the Common Core of Data: School year 2009–10: First look* (Provisional data). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed. gov/pubsearch