Fall 10-23-2009

Who Drops out? A Study of Secondary School Dropouts in Connecticut

Xing Liu
Eastern Connecticut State University, liux@easternct.edu

Y.H. Alison Zhou
Connecticut State Department of Education, alison.zhou@po.state.ct.us

Who Drops Out? A Study of Secondary School Dropouts in Connecticut

Xing Liu

Eastern Connecticut State University

Alison Zhou

Connecticut State Department of Education

Who Drops Out? A Study of Secondary School Dropouts in Connecticut

Introduction

Student dropout has been a great concern of parents, teachers, school administrators and policy makers. It was reported that approximately 5% of students who enrolled in high school in October 2003 dropped out of school within one year (Laird, DeBell, & Chapman, 2006). Those students who dropped out of school might lead to lower salaries and higher unemployment rates (Lehr, 2004).

Researchers, practitioners and policy makers were interested in why students dropped out of school and how to implement feasible strategies to reduce the dropout rate (Allensworth & Easton, 2005, 2007; Dynarski & Gleason, 1999; Gleason & Dysnarski, 2002; Goldschmidt & Wang, 1999; Janosz, LeBlanc, Boulerice, & Tremblay, 2000; Leuchovius, 2006; Smink & Reimer, 2005). A variety of factors which influenced student dropouts were identified and they were complex. Allensworth and Easton (2005) found that two important factors, failures in core courses (number of F's) and the number of full course credits completed during their freshman year, which were identified as the on-track indicator, could predict student dropouts in the ninth grade in Chicago Public Schools. Besides these two factors, their recent study (Allensworth & Easton, 2007) revealed that three additional predictors, student GPA, the number of course failures, and their attendance during the freshman year, could also have played an important role in affecting future student dropouts.

Janosz, LeBlanc, Boulerice and Tremblay (2000) used cluster analysis techniques and the logistic regression model to identify and categorize four types of dropouts: quiet, disengaged,

low-achievers and maladjusted dropouts based on a set of variables, such as, school experience, family experience, peer relationships, leisure activities, beliefs and deviant behaviors. This study suggested that different intervention strategies should be implemented for these four types of dropouts.

However, Gleason and Dysnarski (2002) found controversial results. They investigated the effectiveness of risk factors for identifying students who dropped out of school, and argued that most of the risk factors failed to predict student dropouts. These risk factors included demographic information and family background, previous school performance, school characteristics, personal characteristics and adult responsibilities. Dynarski and Gleason (2002) also examined federally funded dropout-prevention programs, in particular, alternative middle schools for younger students and GED programs for older students. They claimed that understanding the nature of academic, social and personal factors which influenced dropouts and providing particular solutions to these issues would be good strategies to reduce dropout rate.

Identifying risk factors related to dropout was important, however, first of all, it was critical to identify who were most likely to drop out, and what specific characteristics were associated with student dropouts. In a study conducted by Goldschmidt and Wang (1999) using multilevel logistic regression, researchers identified that female students were likely to drop out of middles school than male students. However, Laird, DeBell and Chapman (2006) reported that male students from the ages of 16 to 24 were more likely to drop out of high school than female students. Thus, the results were controversial regarding whether males or females were more likely to be dropouts. In addition, it was uncertain in which grade students were more likely to drop out in secondary schools.

There was a deficiency of research in identifying these characteristics using data from the State of Connecticut. Further, all previous research focused on identifying or predicting the dropouts from a general sample. However, no research on student dropouts has been conducted using the sample of the exited students, who transferred (both in-state and out-of-state), home schooled, graduated, died, or dropped out. Therefore, this study intended to fill the gap and contribute to the research by predicting those dropouts from their characteristics among the exited students.

The purpose of the study was to investigate the relationships between the dropout pattern and student characteristics such as gender, ethnicity and grade level among all exited students from grade 7 to grade 12 in Connecticut. Our research question focused on using demographic information to investigate how the demographic characters of students could accurately predict dropouts among all exited students using logistic regression.

The data collection of tracking student mobility in Connecticut has been evolving after the HR-1, the No Child Left Behind Act of 2001 takes place. The Connecticut State of Education has been collecting dropout data annually since 1991, yet the system of tracking individual student dropout situation begins in 2003. In 2006, the Connecticut local school districts are accountable for all student mobility and mandated to report all exited students.

Methods

Data and Sample

The exited student data were cross-sectional data collected from the Connecticut State Department of Education (CSDE) database. Although longitudinal data tracking student dropout

record could be a much stronger research design for the student dropout study, collecting this kind of data was difficult and it was not readily available. The study included the exited students from grades 7-12 in the 2006-2007 school year. There were 57,709 students. Of these exited students, 3540 (6.13%) were 7^{th} grade students, 6836 (11.85%) were 8^{th} grade students, 6196 (10.74%) were 9^{th} grade students, 4710 (8.16%) were 10^{th} grade students, 6613 (11.46%) were 11^{th} grade students, and 29814 (51.66%) were 12^{th} grade students. The fifty percent of all exited students are from grade 12 because of graduation. There are 27163 (47.07%) were female students, and 30546 (52.93%) were male students.

Variables of interest in the data included student ID, dropout (the dependent variable, indicating whether a student dropped out or not: 1 = dropped out, 0 = not dropped out among exited students), gender (1 = female, 0 = male), and Race/Ethnicity (1 = White American, 2 = Asian American, 3 = Black American, 4 = Indian American, 5 = Hispanic American), Grade (7 = 7^{th} grade, 8 = 8^{th} grade, 9 = 9^{th} grade, 10 = 10^{th} grade, 11 = 11^{th} grade, 12 = 12^{th} grade).

Data Analysis

The survival analysis was inappropriate for this study since the data were not longitudinal. To model students dropped out or not, a logistic regression analysis (Hosmer & Lemeshow, 2000; Long & Freese, 2006) was conducted to examine demographic factors related to secondary students' dropout pattern since the outcome variable was dichotomous. Stata 9.2 was used for analyzing the data. Stata logit command was used for model fitting, and fitstat and listcoef of Stata SPost (Long & Freese, 2006) were used for the analysis of post-estimations for the model. A single explanatory variable model was fitted first, and then a full-model with all the

explanatory variables was fitted. The results of fit statistics, logit coefficients and odds ratios of the independent variables for the fitted models were interpreted and discussed.

Results and Conclusion

Results of the Logistic Regression Analysis with a Single Explanatory Variable

Table 1 presents the results for the logistic regression analysis with a single variable, gender (Model A). The log likelihood ratio Chi-Square test with 1 degree of freedom, LR $\chi^2_{(1)}$ = 38.61, $p < .0001$, indicating that the logit regression coefficient of the predictor, gender was statistically different from 0, so the model with one predictor provided a better fit than the null model with no independent variables in predicting probability for student dropout. The likelihood ratio $R^2_L = .002$, which is the Pseudo R^2, and is also called McFadden's R^2, suggesting that the relationship between student dropout, and the predictor, gender was small.

The estimated logit regression coefficient, $\beta = -.220$, $z = -6.709$, $p < .001$, indicating that gender had a significant effect on student dropout. Odds ratio (OR) = .802, indicating that female students were .802 times the odds for male students of dropping out of school, i.e., female students were less likely than male students to drop out.

Results of the Logistic Regression Analysis for the Full-Model

Next, a logistic regression model with 10 explanatory variables was fitted. This model was referred as the full-model (Model B). Table 1 also provides a summary of the results for the fitting of the full model with 10 explanatory variables. Results indicated that the model with all 10 predictors added significantly improved model prediction, deviance (-2 log likelihood) = 25686.990, the log likelihood ratio Chi-Square test with 10 degree of freedom, LR $\chi^2_{(10)}$ =

3831.98, p < .0001. The likelihood ratio R^2_L = .13, which was much larger than that of the model with a single variable, gender (R^2_L difference = .128), indicating the full-model with 10 predictors provided a better fit than the single-variable model.

In logistic regression, when the explanatory variable was categorical with more than two levels, indicator variables (dummy variables) would be created by setting one of the levels as the reference group. K-1 indicator variables would be needed if the categorical predictor had k levels (Hosmer & Lemeshow, 2000). In this analysis, since ethnicity had five levels, four indicator variables were created internally in stata, and White American were taken as the reference group, thus, odds ratios of the other ethnicity groups compared to the reference category could be estimated. Table 2 presents the specification of the indicator variables for Ethnicity using White as the reference group.

Similarly, five indicator variables were created for Grade with grade 7 as the reference group. Table 3 provides the specification of the indicator variables for Grade using grade 7 as the reference group.

Table 4 provides a summary of estimated logit coefficients and odds ratios for the full-model with 10 predictors. As indicated, compared to female students, male students were more likely to drop out of school (b = -.118, p < .01, OR = .889); compared to White students, both Hispanic students (b = .735, p < .001) and Black students (b = .444, p < .001) were more likely to drop out of school. The odds of Hispanic students dropping out of school were 2.086 times of White students, and the odds of Black students dropping out of school were 1.559 times as high as White students dropping out of school. Compared to students in grade 7, students in grade 10 (b = 1.695, p < .001, OR = 5.445), grade 11 (b = 1.490, p < .001, OR = 4.438), and grade 9 (b =

1.286, $p < .001$, OR = 3.619) were more likely to drop out of school, however, students in grade 12 ($b = -.439$, $p < .001$, OR = .645) and grade 8 ($b = -.377$, $p < .01$, OR = .686) were less likely to drop out of school.

Discussion

This study examined the relationship between dropout patterns and student characteristics such as gender, ethnicity and grade level among all exited students in secondary schools in Connecticut. The current study revealed that gender was a significant predictor of student dropout and male students were more likely to drop out of secondary schools than female students. Our finding supported the previous study by Laird, DeBell and Chapman (2006) which found that male students were more likely to drop out of high school than female students. Results of this study suggested that not only male students from high schools, but also those from middle schools were more likely to drop out, as compared to female students.

Results of the logistic regression models revealed that ethnicity was also a significant predictor of student dropout. Compared to White students, Hispanic students and African American students were more likely to drop out. No significant difference in dropout was identified between White students and Asian American students, or between White students and Native American students. Among these five ethnicity groups, Hispanic students were the most likely to drop out and African American students were the second. Asian American students were the least likely to drop out although there are no significant differences among White American, Native American, and Asian American students. Studies by Allensworth and Easton (2005, 2007) might help to explain why Hispanic and African American students were more likely to drop out since these students were more likely to be low-achievers and have issues with course failure and low attendance.

Grade level was another significant predictor of dropping out. Results indicated that students in grades 10, 11 and grade 9 were more likely to drop out than those in the other grades. Students in grade 12 and grade 8 were less likely to drop out than those in the other four grades. Students from grade 10 were the mostly likely to drop out and students from grade 12 were the least likely to drop out. Students from these grades (10, 11 and 9) were those who were in high schools. A potential reason why students were more likely to drop out from these grades might be that they could not adjust their lives when entering a new school and chose to drop out.

Educational Implications

This study provides empirical evidence of identifying who were more likely to drop out of school among secondary school students in Connecticut. Our results indicated that male Hispanic students were the most likely to drop out, and male Black students were the second most likely to drop out. Our findings also suggested that students in grades 10, 11 and 9 were respectively the top three graders who were most likely to drop out, compared to the other grades in secondary schools. For future research, reasons for dropping out among those high-risk students would be investigated by interviewing teachers and students. Identifying these high risk dropout groups would help practitioners and policy makers develop prevention programs or make interventions to reduce student attrition at the early stage, and thus, to close the achievement gaps among these groups.

References

Allensworth, E. M., & 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: Consortium on Chicago School Research.

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

Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalino, 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.

Barton, P. E. (2002*). Raising achievement and reducing gaps: Reporting progress toward goals for academic achievement in mathematics* (A report to the National Educational Goals Panel).

College Board. (1999). *Reaching the top: A report on the national task force on minority high achievement*. New York: Author.

Dynarski, M., & Gleason, P. (2002). How can we help? What have we learned from evaluation of federal dropout preventive program? *Journal of Education for Student Placed at Risk,* 7(1), 43-69.

Dynarski, M., & Gleason, P. (1999). How can we help? Lessons from federal prevention programs. *Policy Brief.* Princeton, NJ: Mathematics Policy Research, Inc.

Gay, G. (2001). *Improving the achievement of marginalized students of color*: *Including at-risk students in standards-based reform (*A report on McRel's Diversity Roundtable II*)*. Aurora, CO: Mid-continent Research for Education and Learning.

Gleason, P., & Dysnarski, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Student Placed at Risk,* 7(1), 25-41.

Goldschmidt, P., & Wang, J. (1999). When can school affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal, 36(4),* 715-738.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.

Janosz, M., LeBlanc, M., Boulerice, B., &Tremblay, R. E. (2000). Predicting different types of school dropouts: A typological approach with two longitudinal samples. *Journal of Educational Psychology, 92(1),* 171-190.

Laird, J., DeBell, M., Chapman, C. (2006). *Dropout rates in the United States: 2004* (NCES 2007-024). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Lehr, C.A. (2004). Increasing school completion: Learning from research-based practice that work. *Research to Practice Brief: Improving Secondary Education and Transition Services through Research* (ERIC Document Reproduction Service No. Ed 484 287).

Leuchovius, D. (2006). The role of parents in dropout prevention: Strategies that promote graduation and school achievement. *Parent Brief*. National Center on Secondary Education and Transition (NCSET), University of Minnesota (ERIC Document Reproduction Service No. ED 495 870).

Long, J. S. & Freese, J. (2006). Regression models for categorical dependent variables using Stata (2nd ed.). Texas: Stata Press.

No Child Left Behind Act of 2001. (2001). *House Representatives Report 107-334*. U.S. Government Printing Office, Washington, DC.

Smink, J,& Reimer, M. S. (2005). *Fifteen effective strategies for improving student attendance and truancy prevention.* National Dropout Prevention Center Network (ERIC Document Reproduction Service No. ED 485 683).

Steele, C. M. (1999). Thin ice: "Stereotype threat" and black college students. *Atlantic Monthly*, 284(2), 44-54.

Steele, C. M. (1992). Race and the schooling of black Americans. *Atlantic Monthly*, 269(4), 67-78.

Table 1
Model Summaries for the logistic Regression Analysis

	Model A (gender)		Model B (Full-Model)	
Variable	b (se(b))	OR	b (se(b))	OR
Gender [δ]	-.220 (.033)	.802	-.118 (.034)**	.889
Ethnicity_2			-.040 (.121)	.961
Ethnicity_3			.444 (.044)**	1.559
Ethnicity_4			-.274 (.261)	1.315
Ethnicity_5			.735 (.041)**	2.086
Grade_8			-.377 (.110)**	.686
Grade_9			1.286 (.090)**	3.619
Grade_10			1.695 (.090)**	5.445
Grade_11			1.490 (.089)**	4.438
Grade_12			-.439 (.090)**	.645
R^2_L	.002		.130	
Cox & Snell R^2	.001		.064	
McKelvey & Zavoina R^2	.004		.223	
Nagelkerke R^2	.002		.160	
AIC				
BIC				
Model Fit[a] χ^2	38.61 (p < .0001)		3831.98 (p < .0001)	
Deviance	29473.568 (df = 57707)		25686.990 (df = 57698)	

[δ] gender: female=1

[a] Likelihood ratio test

*Significant at p<.05; ** p<.01

Table 2
Specification of the Indicator Variables for Ethnicity Using White as the Reference Group

Ethnicity	Ethnicity_2	Ethnicity_3	Ethnicity_4	Ethnicity_5
White (1)	0	0	0	0
Asian (2)	1	0	0	0
Black (3)	0	1	0	0
Indian (4)	0	0	1	0
Hispanic (5)	0	0	0	1

Table 3
Specification of the Indicator Variables for Grade Using Grade Seven as the Reference Group

Grade	Grade_8	Grade_9	Grade_10	Grade_11	Grade_12
Grade (7)	0	0	0	0	0
Grade (8)	1	0	0	0	0
Grade (9)	0	1	0	0	0
Grade (10)	0	0	1	0	0
Grade (11)	0	0	0	1	0
Grade (12)	0	0	0	0	1

Table 4

Logistic Regression Results for the Full-Model with 10 Explanatory Variables

Variable	b	SE	z	p	OR	95% CI for b	
						Lower	Upper
Gender	-0.118	0.034	-3.45	.001	.889	-0.278	-0.199
Ethnicity_2	-0.040	0.121	-.33	.744	.961	0.359	0.529
Ethnicity_3	0.444	0.044	10.19	.000	1.559	-0.238	0.785
Ethnicity_4	0.274	0.261	1.05	.294	1.315	0.655	0.816
Ethnicity_5	0.735	0.041	18.00	.000	2.086	-0.050	0.974
Grade_8	-0.377	0.110	-3.43	.001	.686	-0.592	-0.161
Grade_9	1.286	0.090	14.32	.000	3.619	1.110	1.462
Grade_10	1.695	0.090	18.83	.000	5.445	1.518	1.871
Grade_11	1.490	0.089	16.73	.000	4.438	1.316	1.665
Grade_12	-0.439	0.090	-4.85	.000	.645	-0.616	-0.262
Constant	-3.346	0.087	-38.67	.000		-3.516	-3.177

Note. b = unstandardized logit coefficient; SE = standard error; z = Wald z-test; p = significant level; OR = odds ratio; CI = confidence interval.