Predicting High School Graduation and Dropout for At-Risk Students: A Multilevel Approach to Measure School Effectiveness

Bidy Raj Subedi, Mark Howard, School District of Palm Beach County, Florida, U.S.A.

Abstract—This paper predicts high school graduation and dropout for at-risk students in one of the largest school districts in the United States using the 2007-2010 Florida high school graduation cohort. We predicted at-risk students’ graduation and dropout statuses employing a two-level hierarchical generalized linear model (HGLM) where the study found the significant effects of student and school level academic and disciplinary factors (such as academic achievement and school suspensions). The school level models predicting student dropout and graduation produced “medium” and “small” effect sizes, respectively.

Index Terms—At-risk students, dropout, graduation, hierarchical generalized linear model, d-type effect size.

I. INTRODUCTION

High school graduation and dropout rates are the key measures of school improvement and educational evaluation. Graduation rates are crucial for high school accountability and are used for making decisions about allocating resources and interventions in low-performing schools. However, dropout is perceived as a reverse of graduation component that warrants several implications for students, teachers, and parents. While controlling the determinants of graduation and dropout rates is crucial for overall school improvement, it is more important for controlling only the significant predictors of these outcome measures in early high school grades. In order to predict student’s graduation and dropout, this study used high school student data from the School District of Palm Beach County (SDPBC), Florida, which is one of the largest urban school districts in the United States with approximately 55,000 students enrolled in grades 9-12 and approximately 173,000 students enrolled in grades K-12.

The at-risk students in Florida high schools are defined as those students who scored Levels 1 and 2 in eighth grade Reading and Mathematics in a standardized test known as Florida Comprehensive Assessment Test (FCAT) for a particular high school graduation cohort. The FCAT is a state mandated standardized criterion-referenced test measuring selected benchmarks in Mathematics, Reading, Writing, and Science which is administered to the students in elementary, middle, and high schools in Florida. In this study for the 2007-2010 high school graduation cohort, the at-risk students are those eight graders who scored Levels 1 and 2 in FCAT Reading and Mathematics in 2006.

A multilevel model, known as hierarchical generalized linear model (HGLM), is employed in order to predict graduation and dropout statuses of at-risk students. Employing this modeling procedure, students’ graduation and dropout events (as binary responses) are predicted incorporating several factors at student and school levels. First we collected the possible predictors at both student and school levels based on past literatures and then identified the strength of these factors in terms of predicting students’ graduation and dropout statuses. We employed a sophisticated statistical model using the real data in which the modeling assumptions and the results of this study were validated.

The study determines the strength of a set of predictors at student and school level models while predicting student graduation and dropout. We have computed the amount of variability within and among schools in their graduation and dropout rates for at-risk students using significant predictors at student and school level models. The effect size at school level is calculated in which a set of statistically significant predictors are incorporated. The effect size would imply the importance of the potential predictors so that proper intervention can be made at student and school levels in order to increase graduation rate and decrease dropout rate. Thus, the central findings of this study can be generalized and implemented in the other school districts having similar demographic characteristics. There are several purposes of this paper. First, school districts in the United States have been searching for an appropriate technique for predicting high school (HS) graduation and dropout based on a theoretical model. Such a model not only can identify significant predictors in estimating at-risk students’ likelihood of HS graduation and dropout, but also can establish a valid and reliable predictive model. Thus, this research explored potential factors affecting HS graduation and dropout based on multilevel technique, called HGLM. Second, this paper determines effect sizes (based on proportion of variance explained) for school level models while predicting HS graduation and dropout events incorporating the best set of
predictors to measure school effectiveness. Thus the implication of this study is that the program planners, teachers, and evaluators at school districts and states will be able to increase graduation and decrease dropout rates for at-risk students by identifying and controlling significant student and school level factors through student intervention programs. Specifically, this study answers the following research questions.

- What are the significant predictors, at student and school levels, of HS graduation and dropouts for at-risk students?
- What are the proportions of variance explained and d-type effect sizes at school level for predicting HS graduation and dropouts?

II. LITERATURE REVIEW

Students’ HS graduation and dropout are important indicators of academic success for the school districts in the United States, and these indicators are even more important for large urban districts. Students attending schools in central cities and larger districts are less likely than students in non-urban and smaller schools systems to graduate [1]. Thus, it is important to develop a predictive model that provides significant predictors of graduation and dropout at student and school levels. Several researchers have predicted high school graduation and academic performance using relevant student and school level factors [2], [3], [4]. For example, researchers not only examined the factors that enabled public school students to significantly improve reading achievement by high school graduation [2] but also they predicted graduation and dropout using significant background and demographic variables [3]. However, some other researchers reported that academic experiences and school factors play a much larger role than student demographics in determining graduation status [4]. Current study uses student academic and disciplinary factors to predict graduation and dropout.

Past studies have predicted high school dropout using relevant predictors in the model [5], [6], [7], [8], [9], [10], [11]. Researchers found that youth who are both antisocial and rejected may be at heightened risk for school dropout [7] and grade retention, school achievement, and school commitment were the best screening variables for potential dropouts [8]. Another study result demonstrates the association of the early home environment, the quality of early care-giving, socioeconomic status, IQ, behavior problems, academic achievement, peer relations, and parent involvement with dropping out of high school at age 19 [9]. The school dropout rate among at-risk students was markedly lower for those students who earlier participated in extracurricular activities compared to those who did not participate [10]. In a recent study, researchers concluded that the decision to drop out is not simply a result of what happens in school, but students’ activities and behaviors outside of school, particularly engaging in deviant and criminal behavior, and (thus) dropping out is more of a process than an event [11].

Past researches address the effect of minority and African-American students [12], [13]. For example, researchers found the student engagement as being an important component of academic resilience among African-American students [13] and some others indicated opportunities regarding early interventions for high school completion for African American students [12]. Researchers also found that males from less poor neighborhoods were more likely to stay in high school [5].

In order to compute the proportion of variance explained and d-type effect size for school level model, we have estimated the variances at both student and school levels. Several studies in past supported the estimation of variance components in binary response model [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Interviewer variability was estimated in a binary response of a problem requiring variance component estimation in a non-normal family [14]. Multilevel model was developed by incorporating an error term in level-1 equation for binary outcomes which can also be derived through a latent variable perspective with the assumption that there exists a latent continuous (outcome) variable (underlying a binary outcome) [19]. Researchers conceived dropping out as an asymptote, or end result, of chronic truancy and estimated variance associated with dropping out in order to measure the effects of high school organization on dropping out [15], and others adopted a general approach for the estimation of variance (at level-1 model) in multilevel nonlinear model using a linearization [13]. Estimates of location and variance parameters were computed for mixed models fitted to binomial data formed by classifying samples from an underlying normal distribution [17] and researches proposed a logistic growth-curve model for the serial dichotomous response with normally distributed random coefficients [15]. Researchers considered a class of probit-normal models for binary data and estimated variance components [22] and random effects model with binary outcome was employed in order to generalize the logistic-normal model for dichotomous response model estimating variance components [23].

A study of graduation and dropout found significant effects of academic and disciplinary factors on graduation and dropout [25]. Although they employed HGLM technique to predict these outcome measures, they did not compute the school level effect sizes in their study, and present study suffices the research needs answering research questions given in the following section. Review of the past researches as well as meaningful correlation between student/school level predictors and high school graduation/dropout suggest using given data with selected predictors in the models employing a two-level HGLM [15], [26], [27], [28]. The advanced statistical analysis technique is used to determine the predictors’ effects, variance components, and d-type effect sizes as suggested by past research [29].
III. METHODS

A. Data Sources, Variables and Model

This study used the data from School District of Palm Beach County (SDBPC) for school year (SY) 2010 high school graduates (cohort 2007-10), provided from Florida Department of Education, which contains three codes: G (graduates), D (dropouts), and N (non-graduates). The dropouts included those students who dropped out in any of the high school grades (9-12). This study included about 2,900 at-risk students from 41 high schools. Table 1 provides a five-year data corresponding to their grades. The layout in Table 1 consists of graduation and dropout data structure from SY2006 through SY2010.

<table>
<thead>
<tr>
<th>School Year</th>
<th>SY 2006</th>
<th>SY 2007</th>
<th>SY 2008</th>
<th>SY 2009</th>
<th>SY 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

Outcome variables

Graduation status: High school graduation status of a student, coded as 1 and 0 for graduation and non-graduation, respectively.

Dropout status: High school dropout status of a student, coded as 1 and 0 for dropout and non-dropout, respectively.

Predictors

Student level predictors

Average achievement: Average of FCAT reading and mathematics development SS scores (a continuous variable).

Average days absent: Average days absent in high school, i.e., total number of days absent in grades 9 through 12 divided by 4 (a continuous variable).

Average days suspension: First calculated the average days of in-school suspension (ISS) and out-of-school suspension (OSS) separately for each student in grades 9 through 12, and then the mean of average days of ISS and OSS was calculated (a continuous variable).

African-American: Student’s race, coded as 1 for African-American student and 0 for others.

English language learner (ELL): Student’s ELL status (i.e., whether or not enrolled in a two-year ESOL program), coded as 1 for ELL students and 0 others.

School level predictors

School mean achievement: The mean of average FCAT reading and mathematics scores for each high school.

School percent retention: Percent of retained (at least one time) students for each high school.

School percent ELL: Percent of ELL students in two-year ESOL program for each high school.

B. Model Development

This study employed a two-level HGLM where separate models were developed and analyzed to predict student graduation and dropout statuses as outcomes. Students’ average achievement, average days suspension, average days absent, ELL status, and African-American (Black) status were used as predictors in level-1 model. School mean achievement, school percent retention, and school percent ELL were used as predictors in level-2 model.

Level-1 and level-2 conditional models (i.e., incorporating predictors in models) were developed based on past literatures as well as the correlations between predictors and outcomes that established the meaningful relationship between above predictors and outcomes. The slopes associated with level-1 and level-2 predictors were estimated in order to answer the first research question. The school-to-school variances were estimated and then effect sizes were computed in order to answer the second research question. The level-2 variance terms can be deleted from the models as they are not significant.

The level-1 conditional model for $i^{th}$ student nested in $j^{th}$ school for predicting graduation is given by $Y_{ij}$. This outcome can be predicted as follows incorporating student level predictors and the random term in level-1 as suggested by researchers [15], [18].

$$\log(P(Y_{ij}=1)/(1-P(Y_{ij}=1))) = \beta_0 + \beta_1 (AVGACH)_{ij} + \beta_2 (AVGSUSP)_{ij} + \beta_3 (AVGSUSP)_{ij} + \beta_4 (ELL)_{ij} + \beta_5 (AFRICANAMERI)_{ij} + e_{ij}$$

(1)

where $\beta_0$ is the intercept. Similarly, the coefficients $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$ are student level slopes or effects for average achievement (AVGACH), average days suspension (AVGSUSP), average days absent (AVGSUSP), student’s ELL status (ELL), and African-American status (AFRICANAMERI), respectively. Further, $e_{ij}$ is student level random term distributed normally with mean zero and constant variance.

Based on the analysis, using only significant school level predictors and residual terms, the level-2 conditional model can be formulated as follows for predicting level-1 coefficients (as outcomes).

$$\beta_0 = \gamma_{00} + \gamma_{01} (SCHLMEANACH)_{j} + \gamma_{02} (SCHLPCTRET)_{j} + u_{0j}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

$$\beta_3 = \gamma_{30}$$

$$\beta_4 = \gamma_{40}$$

$$\beta_5 = \gamma_{50} + \gamma_{51} (SCHLMEANACH)_{j}$$

(2)

It is always useful to formulate a single-equation or combined model in order to mathematically express the individual effects.
and interaction effects of level-1 and level-2 predictors in the same equation. After substituting equation (2) in (1), the single-equation for predicting the probability of graduation can be expressed as follows.

\[
\log(P(Y_{ij} = 1)/(1 - P(Y_{ij} = 1))) = \gamma_0 + \gamma_1 (\text{SCHLMEANACH})_j + \gamma_2 (\text{SCHLPCTRET})_j \\
+ \gamma_3 (\text{AVGACH})_j + \gamma_4 (\text{AVGSUSP})_j + \gamma_5 (\text{AVGABS})_j \\
+ \gamma_6 (\text{ELL})_j + \gamma_7 (\text{AFRICANAMERI})_j \\
+ \gamma_8 (\text{AFRICANAMERI})_j \times (\text{SCHLMEANACH})_j \\
+ e_i + u_{0j} \tag{3}
\]

Equation (3) consists of fixed portion (containing \(\gamma\) terms) and random portion (containing \(e\) and \(u\) terms) of effects, where the term \(\gamma_{0j}\) represents the mean graduation rate for all schools. Further, \(e_i\) represents the random term at student level as defined previously, and \(u_{0j}\) represents the random effect at school level that has multivariate normal distribution. The coefficients \(\gamma_1\) and \(\gamma_2\) are the effects of school mean achievement and school percent of retention, respectively, on average graduation rate. Similarly, the terms \(\gamma_{10}, \gamma_{20}, \text{and } \gamma_{30}\) represent the effects of students’ average achievement, average days suspension, and average days absent, respectively. The term \(\gamma_{40}\) represents the predicted probability of high school graduation for the students with ELL status relative to those students whose status is non-ELL, \(\gamma_{50}\) is the predicted probability of graduation for the group of African-American students relative to the students not included in this group, and \(\gamma_{51}\) is the (interaction) effect of average school achievement for African-American students. The models for predicting dropout status can be formulated and the fixed as well as random coefficients can be defined in a similar manner. Therefore for the sake of conciseness, we do not intend to formulate the equations to predict dropout status.

The fixed effects (intercepts and slopes) and random effect (variance components) at student and school levels are estimated using SAS PROC GLIMMIX procedure [31], [32]. The first research question is answered by estimating fixed effects \(\gamma_s\) and \(p\)-values associated with these effects in Equation (3).

The second research question is answered by estimating the school level variance term \(u_{0j}\) and calculating d-type effect size using the following formula suggested by past research [29].

\[
d = \sqrt{(\text{Variance in achievement lying among school})/\text{(Total student + school variance in student achievement)}} \tag{4}
\]

Since the students are not placed randomly within schools, and student and school level predictors incorporated in (two) separate models provide better estimates of variance and predictors’ effects, our best choice of statistical design to measure the effects of statistical design to measure the effects of school and student level predictors involves selecting the HGLM technique. Researchers suggested that hierarchical or multilevel models can provide a general framework for this type of analysis [30], [27].

We developed two separate models to predict graduation and dropout events employing two-level HGLM incorporating above predictors. For predicting graduation and dropout (statuses), we employed a two-level HGLM with student and school level predictors at level-1 and level-2 models, respectively, where a coefficient in level-1 model is considered as an outcome in level-2 model.

IV. RESULTS

The descriptive statistics associated with predictors at student and school levels are displayed in Table 2.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>N</th>
<th>Mean</th>
<th>SD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average achievement</td>
<td>2877</td>
<td>1772</td>
<td>163.93</td>
</tr>
<tr>
<td>Average days absent</td>
<td>2967</td>
<td>5.54</td>
<td>6.78</td>
</tr>
<tr>
<td>Average days suspension</td>
<td>2967</td>
<td>0.24</td>
<td>0.39</td>
</tr>
<tr>
<td>African-American</td>
<td>2967</td>
<td>0.52</td>
<td>0.50</td>
</tr>
</tbody>
</table>

**“SD” stands for “Standard Deviation”**

Several academic and non-academic predictors are found significant in predicting graduation status. The analysis results in Table 3 provide the list of significant student and school level predictors of high school graduation. The results revealed that average achievement (\(p<.0001\)), average days absent (\(p<.001\)), average days suspension (\(p<.0001\)), and ELL (\(p<.0001\)) are found significant at student level. Similarly, school mean achievement (\(p<.0001\)) and school percent retention (\(p<.0001\)) are found significant at school level. We also examined the interactions among student and school level predictors and found a significant interaction effect (at \(.05\) significance level with \(p=.03\)) between school mean achievement and student’s African-American status. This implies that the school mean achievement for African-American students produced a significant effect on graduation status.

The analysis produced positive effects of academic achievement, at student and school levels, on high school graduation. However, negative effects are found due to average days absent, average days suspension, ELL, school percent retention, and (interaction effect of) school mean achievement for African-American students.
We found significant effects of several academic and non-academic predictors on dropout status. The analysis results presented in Table 4 provide the list of significant academic and non-academic predictors of high school dropout status. The results revealed that average achievement ($p<.0001$) and average days suspension ($p<.001$), with negative and positive effects, respectively, are found significant at student level. The results at school level showed the significant effects of school mean achievement ($p<.0001$) and school percent ELL ($p<.0001$) with negative effects on dropout status. Further, the analysis showed a significant ($p=.035$) positive interaction effect of school mean achievement and student’s African-American status.

**TABLE 3**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Standardized Estimate</th>
<th>S.E.*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average achievement</td>
<td>0.08</td>
<td>0.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Average days absent</td>
<td>-0.03</td>
<td>0.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Average days suspension</td>
<td>-0.07</td>
<td>0.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ELL</td>
<td>-0.23</td>
<td>0.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>School mean achievement</td>
<td>0.14</td>
<td>0.02</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>School percent retention</td>
<td>-0.11</td>
<td>0.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>School mean ach. x African-American</td>
<td>-0.03</td>
<td>0.02</td>
<td>0.030</td>
</tr>
</tbody>
</table>

*“S.E.* stands for “Standard Error”*

We also computed the proportions of variance explained and d-type effect sizes at school level. The results presented in Table 5 display the variance explained at school level for predicting graduation (12%) and dropout (19%), which produced a small (0.35) and medium (0.45) effect sizes, respectively. Note that the effect sizes smaller than 0.39 are considered as “small” and those ranging from 0.39 to 0.45 are considered as “medium” effects by the conventional standards of social science research according to Rowan et al. (2002).

**V. DISCUSSION**

This study predicted graduation and dropout events (as outcomes) employing multilevel model incorporating student and school level predictors and found significant effects on these outcomes due to academic and disciplinary factors. For example, both of these outcomes are significantly predicted due to average achievement (in reading and mathematics) and average days in-school and out of school suspensions (ISS and OSS) at student level and mean achievement at school level. Student status as an ELL enrolled significantly predicted graduation status (negatively). Also, school percent retention and school percent ELL significantly predicted graduation and dropout, respectively. Besides, both graduation and dropout are significantly predicted by school mean achievement for African-American students.

Most of the results of this study are consistent with the past research findings. For example, academic experiences and school factors play a larger role than student demographics in determining graduation rates [4]. Given the significant role of ISS and OSS in students’ activities, the dropout event is a function strongly correlated with student’s misbehavior and misconduct. This fact is supported by past study which mentions that the decision to drop out is not simply a result of what happens in school, but students’ activities and behaviors outside of school, particularly engaging in deviant and criminal behavior [11]. The significant interaction effects between academic achievement and African-American status show that the graduation and dropout events are predicted due to school mean achievement for this (minority) group of students. This can be perceived as “lower will be the graduation rate with a majority of African-American students in a particular school.” It can also be argued that graduation and dropout rates for African-American students do not solely depend on school mean achievement; there could be other possible school level factors that could predict the graduation (positively) and dropout (negatively) for this group of students. A high percentage (67% in SY2010) of African-American students in grades 9-12 participating in free and reduced lunch program in the district [33] and a negative correlation of free and reduced lunch participation with student achievement [25] indicate that there is a different mediating factor besides academic achievement (which possibly is the free and reduced lunch variable) that is causing (decrease in) graduation and (increase in) dropout for African-American students. Further research is necessary to explore such specific factors.
The study found a “medium size” effect for predicting dropout and a “small size” effect for predicting graduation as per the rule to determine the strength of effect size established by past educational research [29]. In addition, this paper developed a method for computing effect sizes for dichotomous outcomes due to variation among schools (in graduation and dropout rates) even though Rowan et al. provided the formula of computing effect sizes only for continuous outcomes. This study extended their approach of computing effect sizes for level-2 model/s for binary outcomes. For this purpose with support of several studies in past, we assumed the level-1 outcome to be approximately normal [20], [34] and computation of level-1 variance in generalized linear model [18], [31], [23]. Thus, the computation of effect sizes due to level-2 (school level) model employing two-level HGLM is a relatively new work of this paper.

The large sample size of the SDBPC, quality data used from authentic sources, high FCAT test reliability (of approximately 0.9), and use of appropriate statistical modeling technique ensured the validity and reliability of the results of this research. The findings of this study will be generalized to the population with similar demographic composition in the United States and other countries.

VI. CONCLUSION

This study identified important predictors at student and school levels employing a HGLM technique predicting high school graduation and dropout for at-risk students in one of the largest school districts in the United States. We have also developed a technique to calculate effect sizes for school level model by incorporating a set of significant predictors in student and school level models. Several technical aspects are discussed in terms of computing effect sizes.

This study brings forth several implications. First, based on the findings of the study, an intervention program is suggested for at-risk students by controlling significant predictors. This would help schools and school districts increase graduation and decrease dropout rates. Second, this study determines effect sizes for school level models (for predicting high school graduation and dropout events) to measure school effectiveness that is useful for research practitioners and educational evaluators. Thus, the teachers, program planners, and evaluators at schools and school districts can increase graduation rate and decrease dropout rate by means of controlling the set of significant predictors based on the strength of effect size. The results of this study can be generalized to other school districts in the nation with similar demographic and racial composition.

There are several limitations of this study. Due to the unavailability of data, this paper could not examine the effects of some additional student level predictors such as students’ extracurricular activities, home environment, and parent involvement. Some other missing school level predictors are principal’s leadership, school climate, school commitment, and percentage highly qualified teachers (with sufficient years of teaching experience, advanced degrees, and teacher certification). Further research is suggested with broader population adding these predictors in a three-level HGLM considering student, teacher, and school data at level-1, level-2, and level-3 models, respectively.

REFERENCES


Advances in Education                  Vol.2  No.1         Feb.  2013      ISSN 2165 - 946X

Bidya R. Subedi received his Ph.D. in Educational Measurement and Statistics from Florida State University and master degree in Educational Psychology from Michigan State University. He has been working as a Specialist in Assessment & Technology in the Department of Research, Evaluation, and Assessment, School District of Palm Beach County, Florida, USA. His research areas include hierarchical linear and nonlinear modeling, high school graduation and dropouts, student achievement and gains in K-12 as well as teacher and school effectiveness.

Mark Howard received his master degree in Educational Technology from Ball State University. He is the Director of the Department of Research, Evaluation and Assessment in School District of Palm Beach County, Florida, USA. His research interests include student achievement, high school graduation and dropout as well as teacher and school evaluations.