

Analysis of UA First-time Freshmen Success in English 111 by Course Delivery Type

Chantelle J. McGinness
University of Alaska Fairbanks
Spring 2013

Committee:
Dr. Ronald Barry (Advisor)
Dr. Julie McIntyre
Dr. Margaret Short

Author Note

Chantelle McGinness, Department of Mathematics and Statistics, University of Alaska Fairbanks.
Chantelle McGinness is also employed by UA Institutional Research and Analysis, University of Alaska.

This research was preformed to satisfy degree requirements for a Master of Science in Statistics at the University of Alaska Fairbanks.

Correspondence concerning this article should be addressed to Chantelle McGinness, UA Institutional Research and Analysis, University of Alaska, PO Box 755260, Fairbanks, AK 99775.
E-mail: cjmcginness@alaska.edu

TABLE OF CONTENTS

	<u>PAGE</u>
ABSTRACT	3
1. Introduction.....	4
2. Empirical Data.....	6
3. Logistic Regression.....	8
4. Classification and Regression Tree (CART) Analysis.....	19
5. Discussion.....	26
REFERENCES	30
APPENDICES	
Appendix A.....	32
Appendix B	34
Appendix C	37

Abstract

This study examines course success for degree-seeking first-time freshmen who enrolled in English 111 during all terms from fall 2011 to fall 2013. Two data sets were analyzed, one for students who took the course by e-Learning and the other for students who took the course by traditional face-to-face meetings. The response variable was student success in the course, i.e. a student earned a C or better. Explanatory variables that were investigated included term, gender, race, origin at entry, age, first-generation status, financial aid status, high school grade point average, degree level, full-time/part-time status, student credit hour load, and broad degree discipline area. For each data set, logistic regression was performed, followed by classification and regression tree (CART) analysis, to determine which variables were associated with student success. Similarities and differences in the models that resulted for each course delivery type (e-Learning or traditional) were discussed. Finally, the results of the logistic regression and the CART analysis were also compared and contrasted.

Keywords: e-Learning, post-secondary education, online learning, logistic regression, classification and regression tree analysis

1. Introduction

In 2009 the Alaska State Legislature published an audit report addressing University of Alaska's utilization of e-Learning services, encouraging institutions across the University of Alaska System to develop a more collaborative and student-centric approach to delivering e-Learning courses (Davidson, 2009). As a result, in FY10 University of Alaska adopted the following definition for e-Learning:

Planned learning that predominantly occurs in situations where a student is not required to be in a predetermined location. As a result, e-Learning courses require a different course design and development, different pedagogical techniques, and communication through instructional technologies. (University of Alaska, 2012).

This new definition was integrated into University of Alaska's Banner Information system in FY11, expanding upon the old criteria used to identify e-Learning courses (University of Alaska, 2012). University of Alaska's e-Learning course offerings range across preparatory to undergraduate, professional, and graduate levels (University of Alaska, 2012). Faculty may use any combination of one or more of the following technologies to deliver e-Learning courses: audio or video conferencing, correspondence, internet, live television, or multimedia (University of Alaska, 2011).

Nationally, the percentage of undergraduate students enrolled in at least one distance education course increased from 8 percent in 2000 to 20 percent in 2008 (Radford, 2011). The National Center for Education Statistics (NCES) reported that 21 percent of undergraduate students attending public institutions enrolled in an e-Learning course in FY08 (NCES, 2010). Comparatively, nearly 29 percent of University of Alaska's undergraduate students took at least one e-Learning course in fiscal year 2008 (FY08), rising to 34 percent by FY12.

Considering other emerging trends in e-Learning, Lee and Choi (2011) preformed a literature review of recent research that investigated factors affecting online course dropouts, restricting the review to peer-reviewed journals that were published between 1999 and 2009. Of the 159 studies originally identified, Lee and Choi (2011) synthesized 35 empirical studies, identifying 44 factors associated with dropout rates of online courses. Definitions for e-Learning varied among the studies, with some studies failing to provide any definition at all. Academic factors determined to be associated with student success in e-Learning courses included grade point average and SAT math score (Lee and Choi, 2011). Generally, students with lower performance scores succeeded in e-Learning courses at a lower rate (Lee and Choi, 2011). Other factors associated with student success in e-Learning courses were work commitments, financial aid, and life circumstances (Lee and Choi, 2011). Demographic characteristics were not included in the significant factors identified “because the findings of many studies were incompatible with one another regarding the relationship between demographics and student’s persistence in online courses” (Lee and Choi, 2011). However, in a recent study, Xu and Jaggars (2013) investigated how adaptability to online learning differed across student subgroups, reporting that males, younger students, black students, and students with lower grade point averages were not as successful as others in online courses, especially English and Social Sciences courses. It is important to note that generalizability of research findings to University of Alaska is questionable, considering differences in the student populations and online course structures studied.

This study focused on English 111 courses delivered at the University of Alaska from fall 2011 through fall 2012. The primary purpose of this study was to identify significant demographic and academic factors associated with course success for degree-seeking first-time freshmen in English 111 delivered via e-Learning. A corresponding analysis was performed for first-time freshmen who took

the course via traditional face-to-face meetings, in order to compare whether the factors that were associated with course success for e-Learning students were the same for traditional students. For each course delivery method, two statistical models were developed using two different statistical methods, namely logistic regression and classification and regression tree (CART) analysis, and the results of each model were compared and contrasted.

2. Empirical Data

2.1 Data

The data for this study was extracted from the University of Alaska Decision Support Database (DSD) Closing Extracts using a SAS/SQL query. DSD Closing extracts are simply snapshots of each student's academic activity at a certain (close) date for each term.

Data was extracted for every degree-seeking, first-time freshman enrolled at the University of Alaska enrolled in English 111 course across the University of Alaska System from fall 2011 through fall 2012. Earlier data was not used because the new definition for e-Learning became effective in fall 2011. Only courses that were identified as GPA-eligible were considered, which excluded from the analysis students who audited the course, withdrew from the course, or received an incomplete in the course. Since these students didn't receive a grade in the course, success or failure could not be determined for them.

Course delivery type is based on the percentage of location-based time for which students received instruction. Traditional face-to-face meetings indicate that students received instruction for at least 51 percent of total contact hours with the instructor at a defined location, whereas e-Learning courses were delivered independent of location for at least 50 percent of the total contact hours.

2.2 Variables

The response variable was course success, which was a binary variable identifying whether a student passed the course. A student was successful if the student received one of the following grades: A+, A, A-, B+, B, B-, C+, and C.

A number of demographic variables were considered as predictors. These included the following:

1. Gender: Gender is self-reported to UA on the application for admissions.
2. UAR Race: UAR Race is the race identified for each student for reporting purposes. Race is self-reported on the application for admissions, and each student may pick one or all of the 18 race options. In order to generate unduplicated headcounts for race, University of Alaska conventionally identifies a primary race classification for each student, referred to here as UAR race, which include the following categories: Alaska Native/American Indian, Hawaiian Native/Pacific Islander, Asian, Black, White, and Not Reported.
3. Origin at Entry Status: Origin at entry status is a flag that identifies a student's origin at entry: Alaska, Out-of-state, or International. Origin at entry is the location of a student when first enrolling at the university and does not reflect a student's official residency status.
4. Age: Age is how old a student is on October 1 of the cohort year. Each student was assigned to one of the following age categories: Under 20, 20-24, 25-29, 30-39, Over 40.
5. First Generation Status: First-generation status is self-reported on the application for admissions and on the Free Application for Federal Student Aid. A student is considered first-generation if the student's parents' highest education completed was high school-level or lower.
6. Financial Aid: Financial aid status is a flag that identifies whether the student received financial aid during the enrollment term.

Several academic variables were also used as predictors, and included the following:

1. Term: This variable identifies which term a student was enrolled in English 111: fall 2011, spring 2012, summer 2012, or fall 2012.
2. High School Grade Point Average: This variable is a continuous variable on a 0.0 to 5.0 scale.
3. Degree Level: Degree level is the primary degree program in which a student is enrolled. Categories include certificate, associate, or bachelor.
4. Full-time/part-time Status: Full-time/part-time status is determined each term by a student's credit hour load. Students who were enrolled in 12 or more credits were classified as full-time students. Students who were enrolled for fewer credits were classified as part-time students. Audited student credit hours were not included for computing full-time status.
5. Student Credit Hour Load: Student credit hour load is based on the number on non-audit hours taken by students in credit courses: 0-3, 4-6, 7-8, 9, 10-11, 12-14, 15 or more.
6. Discipline Area: Discipline area is the student's broad field of study, based on the first two digits of discipline (CIP) code for a student's primary major. Examples include education, engineering, health, natural resources, etc.

3. Logistic Regression

3.1 Method of Analysis

The logistic regressions for this study were generated using SAS software, Version 9 of the SAS System for Linux. Copyright © 2002-2010 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Developed during the late 19th century by Sir Francis Galton, regression analysis is a statistical method used to investigate how a response variable varies with one or more predictor variables (Kutner et al., 2005). The models that result from regression analysis are widely used for prediction (Kutner et al., 2005). Regression models may be linear or nonlinear and appear in the general form

$$Y_i = f(\mathbf{X}_i, \boldsymbol{\beta}) + \varepsilon_i,$$

where Y_i is the value of the response variable for the i^{th} observation, $f(\mathbf{X}_i, \boldsymbol{\beta})$ is the response function, $\mathbf{X}_i = [X_{1i}, \dots, X_{pi}]'$ is vector of the observations on the predictor variables for the i^{th} observation, $\boldsymbol{\beta} = [\alpha, \beta_1, \dots, \beta_p]$ is the vector of the regression coefficients, and ε_i is the random error associated with the i^{th} observation (Kutner et al., 2005). The error terms are assumed to be uncorrelated with mean zero and constant variance (Kutner et al., 2005). The regression analysis serves to estimate the regression coefficients, which in turn, collectively indicate how the probability distribution of Y behaves in response to the predictor variables X_1, \dots, X_p (Kutner et al., 2005). Linear regression models occur when the parameters all appear in the first-order in the response function $f(\mathbf{X}_i, \boldsymbol{\beta})$ (Kutner et al., 2005). The following models are all examples of linear regression models:

$$Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

$$Y_i = \alpha + \beta_1 (X_{1i})^2 + \beta_2 \exp(X_{2i}) + \varepsilon_i$$

$$\log(Y_i) = \alpha + \beta_1 (X_{1i})^2 + \beta_2 \exp(X_{2i}) + \beta_3 (X_{1i} X_{2i}) + \varepsilon_i$$

Nonlinear regression models occur when the response function $f(\mathbf{X}_i, \boldsymbol{\beta})$ is nonlinear in the parameters (Kutner et al., 2005). For example, the following models are all considered nonlinear regression models:

$$Y_i = \alpha \exp(\beta_1 X_{1i}) + \varepsilon_i$$

$$Y_i = \alpha + \beta_1 (X_{1i})^{\beta_2} + \varepsilon_i$$

$$Y_i = \alpha / \exp(\beta_1 X_{1i}) + \beta_2 (X_{2i})^{\beta_3} + \varepsilon_i$$

Logistic regression is a specific form of regression in which the response variable Y is a Bernoulli random variable with the parameter $\pi_i = P(Y=1 | X_{1i}, \dots, X_{pi})$, where

$$\pi_i = \exp(\alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi}) / (1 + \exp(\alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi})),$$

and thus, the nonlinear logistic regression function would take the form:

$$Y_i = \pi_i + \varepsilon_i$$

(Kutner et al., 2005). Transformations may be applied to nonlinear response functions, allowing them to become linearized (Kutner et al., 2005). The transformation used in logistic regression is called the 'logit' and is defined to be

$$\text{logit}[\pi_i] = \log(\pi_i / (1 - \pi_i)) = \alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi},$$

which equates the logit response function to a linear combination of the predictors (Kutner et al., 2005).

With respect to parameter interpretation in logistic regression, the intercept parameter α is not typically considered (Agresti, 2002). The beta parameters β_1, \dots, β_p , however, are important. The sign of each beta parameter indicates whether π_i increases or decreases as the associated predictor variable increases (Agresti, 2002). Exponentiating a particular β_j results in the interpretation that, for a one-unit increase in the predictor variable X_j that is associated with β_j , the odds of success increase multiplicatively by $\exp(\beta_j)$ (Agresti, 2002).

Logistic regression was used to determine whether course delivery method was statistically related to course success at a 5% significance level. Afterwards, the data was divided into two subsets, one for each delivery method, and forward stepwise logistic regression was used to determine which demographic and academic variables were statistically related to course success for each delivery method at a 5% significance level.

Main effects models for each data set were built using forward stepwise selection. This procedure first estimated a parameter for the intercept model, then evaluated the explanatory variables using the residual chi-square score statistic at a 5% significance level (SAS Institute Inc.,1999). If the explanatory variable with the highest residual chi-square score statistic was significant, that variable was added to the model (SAS Institute Inc.,1999). Then the effects included in the model were tested again at a 5% significance level by backwards elimination using the Wald chi-squared test statistic, which was computed by dividing an estimate by its estimated standard error and squaring that quantity (SAS Institute Inc.,1999). Alternating between forward selection and backward elimination is repeated until no explanatory variables met the 5% significance criteria to enter in the model or on the final backward elimination step, the only explanatory variable removed was the one that was added to the model as a result of the most recent forward selection step (SAS Institute Inc.,1999).

The optimization technique used for generating maximum likelihood estimates was the Fisher-scoring algorithm. Wald confidence intervals were estimated for odds ratios. Goodness of fit for models resulting from forward stepwise selection was determined by Hosmer and Lemeshow tests, which used a Pearson's chi-squared test statistic to test the null hypothesis that the models fit well. Regarding regression diagnostics, Pearson residuals were used to detect observations that are poorly described by the model, hat matrix diagonals were used to detect extreme points in the design space, and DFBETAS statistics were used to assess the effect of each observation on each parameter estimated by the model (SAS Institute Inc., 2008). For a thorough explanation of these regression diagnostics, readers are referred to pages 4,132-4,134 of the SAS 9.3 User's Guide (20). To measure predictive power of the models, receiver operating characteristic curves (ROC) were analyzed.

3.2 Results: Course Delivery Method

The initial logistic regression investigated whether course success was statistically related to course delivery method, traditional versus e-Learning. In total, 2,255 first-time freshmen were identified as having taken English 111 between fall 2011 and fall 2012. After 207 records were excluded from the analysis because of missing high school GPA, 2,048 records were analyzed. Table 1 reports the demographic characteristics by course delivery type.

Table 1. Demographic Characteristics by Course Delivery Type

	eLearning		Traditional	
	Headcount	Percent of Total	Headcount	Percent of Total
Gender				
Female	125	69.44	1029	55.09
Male	55	30.56	839	44.91
UAR Race				
AK Native/Amer. Indian	16	8.89	217	11.62
Hawaiian/Pacific Islander	3	1.67	20	1.07
Asian	12	6.67	102	5.46
Black	3	1.67	65	3.48
White	99	55.00	1012	54.18
Other	47	26.11	452	24.20
Origin at Entry				
Alaskan	173	96.11	1713	91.70
Out-of-State	7	3.89	17	0.91
Foreign			138	7.39
Age Group				
Under 20	94	52.22	1598	85.55
20-24	46	25.56	161	8.62
25-29	22	12.22	73	3.91
30-39	10	5.56	22	1.18
40 or Older	8	4.44	14	0.75
First Generation Student				
No	123	68.33	1278	68.42
Yes	57	31.67	590	31.58
Received Financial Aid				
No	112	62.22	1060	56.75
Yes	68	37.78	808	43.25
Total	180	100.00	1868	100.00

Table 2 shows results for the logistic regression based on the following model:

$$\text{logit}[\pi_i] = \alpha + \beta_1 X_{1i}, \quad (3.1)$$

where $\pi = P(\text{success}=1)$, α and β_1 are the regression parameters, and $X_{1i} = 1$ for e-Learning course delivery method and 0 for traditional course delivery method.

Table 2: Analysis of Maximum Likelihood Estimates for Model with Course Delivery Method

Parameter	df	Estimate	Standard Error	Wald Chi-square	p-value
Intercept	1	1.6538	0.063	688.7199	<.0001
eLearning	1	-0.8328	0.1736	23.0013	<.0001

The final model used to estimate the probability of success for first-time freshmen in English 111 based solely on course delivery method was:

$$\text{logit}[\pi_{ihat}] = 1.6538 - 0.8328X_{1i},$$

where $X_{1i} = 1$ if a student took the course via e-Learning and 0 if the student took the course via traditional face-to-face meetings.

The following null hypothesis was tested to determine whether e-Learning course delivery method contributed significantly to the model

$$H_0: \beta_1 = 0.$$

Comparing the maximized log-likelihood L_I for the full model (3.1) to L_0 for the simpler intercept only model, the likelihood ratio statistic $-2(L_0 - L_I) = 20.9448$ has $df=1$, the difference between the number of parameters in each model. With chi-squared $p<0.0001$, e-Learning course delivery method was statistically significant. Since the explanatory variable is categorical, this is equivalent to a chi-square test of independence of course delivery versus course success.

As supported by Table 3, first-time freshmen who took English 111 via e-Learning were 0.435 as likely, or equivalently, 2.3 times less likely, to succeed in the course than first-time freshmen who took the course via traditional face-to-face meetings.

Table 3: Odds Ratio Estimates for Course Delivery Method

Effect	Estimate	95% Wald	
		Confidence Limits	
eLearning vs Traditional face-to-face	0.435	0.309	0.611

3.3 Results: Courses Delivered via e-Learning

For students who took English 111 via e-Learning, 40 records were excluded from the analysis because of missing high school GPA. Therefore, 180 observations were used for the analysis.

The forward stepwise logistic regression procedure retained, in order of entry to the model, high school GPA and financial aid status. Table 4 summarizes the stepwise selection procedure.

Table 4: Summary of Stepwise Selection Modeling Course Success in English 111 via eLearning

Step	Effect		df	Number In	Score	
	Entered	Removed			Chi-Square	p-value
1	High School GPA		1	1	12.7557	0.0004
2	Financial Aid Flag = no		1	2	4.4347	0.0352

Table 5 shows parameter estimation resulting from the forward stepwise logistic regression procedure based on the initial intercept model:

$$\text{logit}[\pi] = \alpha \quad (3.2)$$

where $\pi = P(\text{success}=1)$, and α is the intercept parameter.

Table 5: Analysis of Maximum Likelihood Estimates Modeling Course Success in English 111 via eLearning

Parameter	df	Estimate	Standard Error	Wald Chi-square	p-value
Intercept	1	-2.2517	1.0126	4.9451	0.0262
High School GPA	1	1.2005	0.3399	12.4744	0.0004
Financial Aid Flag = no	1	-0.7679	0.3685	4.3435	0.0372

The final model used to estimate the probability of success for first-time freshmen in English 111 via e-Learning was:

$$\text{logit}[\pi_{\text{hat}}] = -2.2517 - 0.7679 X_{1i} + 1.2005 X_{2i},$$

where $X_{1i}=1$ if a student did not receive financial aid and 0 if a student did receive financial aid, and X_{2i} =high school GPA.

The Hosmer and Lemeshow test supported goodness of fit $p=0.7837$, as shown in Table 7, suggesting that there is not a significantly better model based on a nonlinear function of high school GPA and financial aid.

For a students who took the English 111 via e-Learning, regardless of financial aid status, an one unit increase in high school GPA had a multiplicative effect of 3.3 on the odds that a student succeed in the course. Holding high school GPA constant, students who did not receive financial aid were 0.464 times as likely to succeed as students who did receive financial aid or said alternatively, students who received financial aid were 2.2 times more likely succeed in the course. These results are summarized in Table 6, which includes 95% Wald confidence limits.

Table 6: Odds Ratio Estimates for Significant Effects for eLearning Model

Effect	Estimate	95% Wald	
		Confidence Limits	
High school GPA	3.322	1.706	6.467
Financial aid no vs yes	0.464	0.225	0.955

Using standard cutoff values of ± 2 , the Pearson residuals and the deviance residuals indicated that the model poorly accounted for three observations. Based on a cutoff value of $2p/n=0.0222$, where $p=2$ for the number of parameter in the model and $n=180$ for the number of observations in the model, evaluation of the hat matrix diagonals suggested that there were 28 extreme points in the design space. The DFBETAS statistics suggested that there were no cases that caused instability in estimating the parameters.

Measuring predictive power of the model, the area under the ROC curve, also known as the concordance index, was 0.664 for the model containing only high school GPA and 0.689 for the

model including high school GPA and financial aid status, which is good (anything above 0.5 implies prediction better than random guessing).

3.4 Results: Courses Delivered via Traditional Face-to-Face Meetings

For students who took English 111 via traditional face-to-face meetings, 167 records were excluded from the analysis because of missing high school GPA, resulting in 1,868 observations being used for the analysis.

To investigate variables associated with course success for students who took English 111 via traditional face-to-face meetings, forward stepwise logistic regression was performed based on the same initial intercept model as model (3.2) and results are summarized in Table 7. Of all the explanatory variables tested, the forward stepwise logistic regression procedure retained, in order of entry to the model, high school GPA, age group, origin at entry, and full-time/part-time status.

Table 7: Summary of Stepwise Selection Modeling Course Success in English 111 via Traditional Face-to-face Meetings

Step	Effect		df	Number In	Score	
	Entered	Removed			Chi-Square	p-value
1	High School GPA		1	1	158.189	<.0001
2	Age Group		4	2	24.3703	<.0001
3	Origin at Entry		2	3	9.5395	0.0085
4	Full-time/Part-time Status		1	4	3.9826	0.046

As supported by Table 8, the final model used to estimate the probability of success for first-time freshmen in English 111 via e-Learning was:

$$\begin{aligned} \text{logit}[\pi_{\text{ihat}}] = & -3.66 + 1.61 X_{1i} + 0.44 X_{2i} + 1.05 X_{3i} + 15.65 X_{4i} + 2.49 X_{5i} \\ & + 0.42 X_{6i} + 3.54 X_{7i} + 0.36 X_{8i} \end{aligned}$$

where d=high school GPA, X_{2i}, \dots, X_{5i} are indicator variables for age groups 20-24, 25-29, 30-39, 40 or older, respectively, X_{6i} and X_{7i} are indicator variables for out of state and foreign origins at entry, respectively, and X_{8i} identifies whether a student was enrolled full-time. The p-value ($p=0.9774$) for

Age Group 30-39 suggests that students in this age group did not perform significantly different than students aged 20 and younger, however the estimate and the associated standard error are very large. This resulted from quasicomplete separation of this variable. Of the 22 first-time freshmen who were aged 30-39, all succeeded in the course. This quasicomplete separation does not invalidate the model, but rather the model perfectly predicts course success for first-time freshmen aged 30-39. With $p=0.1249$, first-time freshmen whose origin at entry was out of state did not perform significantly different than Alaskan students. However, evidence suggests that foreign students were significantly more likely to succeed in the English 111 delivered traditionally than Alaskan students ($p=0.0075$).

Table 8: Analysis of Maximum Likelihood Estimates Modeling Course Success in English 111 via Traditional Face-to-face Meetings

Parameter	df	Estimate	Standard Error	Wald Chi-square	p-value
Intercept	1	-3.6605	0.4129	78.5819	<.0001
High School GPA	1	1.6141	0.134	145.0384	<.0001
Age Group 20-24	1	0.439	0.2301	3.6409	0.0564
Age Group 25-29	1	1.05	0.3526	8.8663	0.0029
Age Group 30-39	1	15.6493	551.2	0.0008	0.9774
Age Group 40 or Older	1	2.4884	1.083	5.2795	0.0216
Origin at Entry Out of State	1	0.4237	0.2761	7.1433	0.1249
Origin at Entry Foreign	1	3.5386	1.324	2.3552	0.0075
Full-time/Part-time Status = full-time	1	0.3608	0.1812	3.9628	0.0465

* Baselines: Origin at Entry Alaska Age Group Under 20, full-time/part-time status = part-time

The Hosmer and Lemeshow test supported goodness of fit with $p=0.76$, as shown in Table 11.

Table 9 summarizes the odds ratios associated with each parameter in the final model. Holding all other variables constant, for one unit increase in high school GPA, first-time freshmen who took English 111 traditionally are 5.0 times more likely to succeed. With respect to age, students who were under 20 were least likely to succeed in English 111 delivered via traditional face-to-face meetings, with student aged 20-24 being 1.5 times more likely to succeed, those aged 25-29 being 2.8 times more likely to succeed, and students 40 and older being 12.0 times more like to succeed than

student aged 20 and younger. Considering student's origin at entry status, students who had a foreign origin at entry were 3.5 times more likely to succeed in the course than Alaskan students. Students who were enrolled full-time were 0.36 times more likely to succeed than part-time students.

Table 9: Odds Ratio Estimates for Significant Effects for Traditional Model

Effect	Estimate	95% Wald	
		Confidence Limits	
High School GPA	5.024	3.863	6.533
Age Group 20-24 vs Under 20	1.551	0.988	2.435
Age Group 25-29 vs Under 20	2.858	1.432	5.704
Age Group 30-39 vs Under 20	>999.999	<0.001	>999.999
Age Group 40 or Older vs Under 20	12.042	1.442	100.584
Origin at Entry Out of State vs Alaska	34.418	2.569	461.053
Origin at Entry Foreign vs Alaska	1.528	0.889	2.625
Full-time vs Part-time	1.434	3.863	6.533

Using standard cutoff values of ± 2 , the Pearson residuals and the deviance residuals indicated that the model poorly accounted for 72 observations. Based on a cutoff value of $2p/n=0.0086$ where $p=2$ for the number of parameter in the model and $n=1868$ for the number of observations in the model, evaluation of the hat matrix diagonals suggested that there were 227 extreme points in the design space. The DFBETAS statistics suggested that there were no cases that caused instability in estimating the parameters.

Measuring predictive power of the model pertaining to English 111 via traditional face-to-face meetings, the concordance index was 0.735 for the model containing only high school GPA. When age group was added to the model with high school GPA, the concordance index increased to 0.749. Including the origin at entry status along with high school GPA and age group again increased the concordance index to 0.753. Finally, considering the model including all four variables that were identified as significant in the stepwise logistic regression procedure resulted in a concordance index of 0.755.

4. Classification and Regression Tree (CART) Analysis

4.1 Method of Analysis

Prediction trees are nonparametric models that are useful for easily predicting a response variable from a set of observed explanatory variables. Buja and Lee (2001) summarize tree construction as beginning with “a greedy growing phase driven by a binary splitting criterion, followed by a pruning phase based on cost-complexity measures and/or estimates of generalization error.” Response variables are categorical for classification trees and continuous for regression trees (Blasius and Greenacre, 1998).

Beginning at the “root” or top of the tree, all possible binary splits of every explanatory variable are considered, and the best split is identified by a statistical criterion, which serves as a measure of impurity. The objective of the split is to divide the data into two groups such that the observations within each group are as homogeneous as possible (Blasius and Greenacre, 1998). The split that maximizes the reduction in impurity is chosen, resulting in two daughter nodes (Blasius and Greenacre, 1998). Then the splitting process repeats at each node and continues to split the resulting nodes until a stopping criterion is met (Blasius and Greenacre, 1998).

One disadvantage of tree models is that they tend to overfit the data in that the model perfectly “predicts” observations in the data set used to build the tree, but poorly predicts new observations. It is possible for splitting to continue until there is a terminal node, or leaf, for every observation in the data set. The “greedy growing phase” typically results in a tree with many more nodes than is optimal. Therefore, once the initial tree is developed, cross-validation is applied to determine how complex the tree should be based on the minimization of the cost-complexity measure,

$$R_{\alpha}(T) = R(T) + \alpha \text{ size}(T),$$

where $R(T)$ is a measure of lack of fit, which was deviance in this study, and $\alpha > 0$ penalizes for tree size (Ripley, 1996).

The CART analyses for this study were generated using R software, Version 2.15.2, on Windows 7. R software is free software under the terms of the GNU General Public License as published by the Free Software Foundation.

The 'tree' package was used to generate the classification trees. The splitting criterion used in this study was deviance, which is defined to be

$$D(T) = 2[\sum_t n_t \log n_t - \sum_c n_{tc} \log n_{tc}],$$

where t is the leaf index and c is the class index, n_t is the number of observations that will reach leaf t , and n_{tc} is the number of each class at the node (Ripley, 1996). Using deviance as the splitting criterion results in an impurity index of

$$I(T) = D(T)/2n,$$

where n is the total number of observations (Ripley, 1996). For each data set, an initial tree was developed using the `tree()` function with a stopping criterion of a minimum within-node deviance of $0.003 \cdot D_r$, where D_r is the deviance of the root node. Then 10-fold cross-validation was performed to determine the optimal size of the tree, using the function `cv.tree()`. Finally, the best tree having the number of terminal nodes identified by the cross validation was generated.

4.2 Results: Courses Delivered via e-Learning

Using all explanatory variables, an initial classification tree was generated for students who took English 111 via e-Learning, which is depicted in Figure 1. This tree was rather complex, having 22 terminal nodes and included the following variables: discipline, high school GPA, student credit hour load, term, age group, race, gender, and degree level. The misclassification error rate for this tree was $0.1444 = 26 / 180$. The cross-validation procedure did not provide useful results, as evident by the

cross-validated deviance increasing for splits starting as early as the root node in Figure 2. Therefore, a final best tree was not generated for this data set.

In predicting course success, interpretation of a classification tree begins at the root node. In Figure 2, the root node resulted in an initial binary split on the broad degree discipline. Appendix A contains the diagram details for this tree. If a student's broad degree discipline belongs to one of the following categories, 01 Business and Public Administration, 02 Computer and Information Science, 03 Education, 04 Engineering, 06 Health, 07 Letters, Comm., and Philosophy, 10 Social Sciences, 11 Visual and Performing Arts, or 12 Vocational Education, the left split is chosen, and the variable at the next node will be considered. By choosing the left or right split at each node, progression down the tree continues until a terminal node is reached. For example, if a student has a broad degree discipline other than one listed above, the right split of the root node is evaluated, which is a terminal node that predicts the student will succeed in the course.

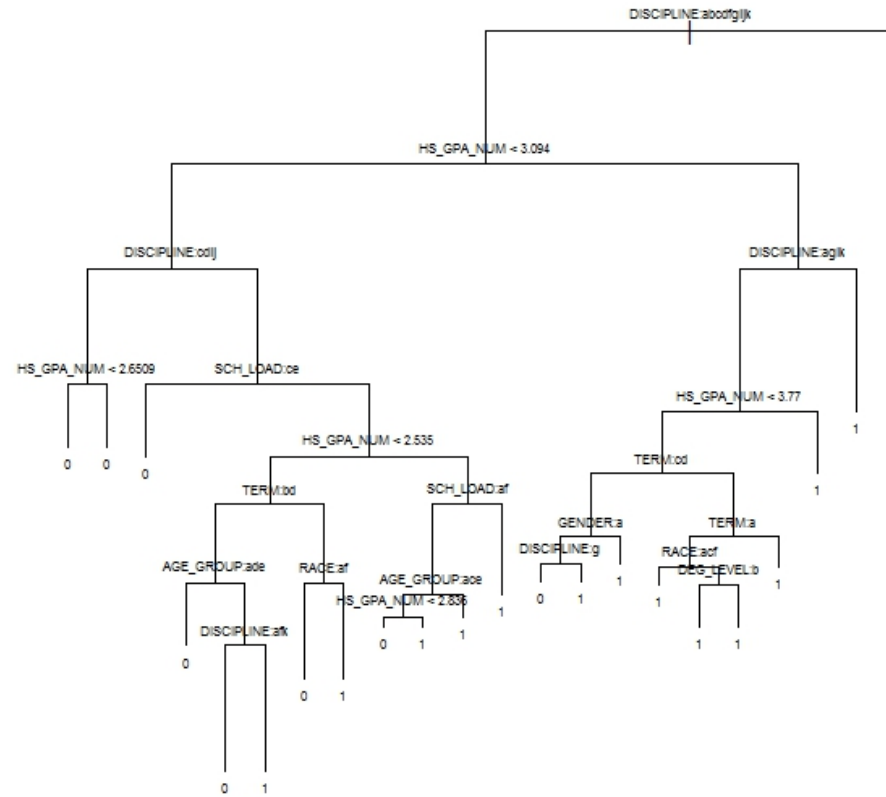


Figure 1: Initial classification tree for student who took English 111 via e-Learning (Diagram details can be found in Appendix A).

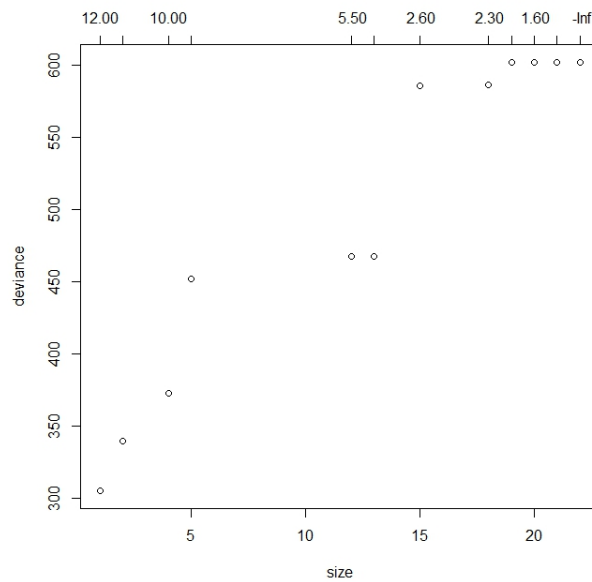


Figure 2: Tree size (horizontal axis) versus cross-validated deviance (vertical axis) for successive prunings of the initial classification tree for student who took English 111 via e-Learning. (The upper scale on the horizontal axis refers to the “cost/complexity” penalty.)

4.3 Results: Courses Delivered via Traditional Face-to-face Meetings

For students who took English 111 via traditional face-to-face meetings, all explanatory variables were used to construct the initial classification tree, which is shown in Figure 3. This initial classification tree had 41 terminal nodes and included the following variables: high school GPA, age group, student credit hour load, term, discipline, gender, race, and financial aid status. The misclassification error rate for this tree was $0.1383 = 258 / 1,868$. As shown in Figure 4, the cross-validation procedure indicated that the best tree has 5 terminal nodes. The initial tree was pruned to generate the best tree with 5 terminal nodes. The result is shown in Figure 5. The final tree allows predictions based on high school GPA and age group. However, this final tree does not appear to be useful for predicting course success because all terminal nodes predict that a student will succeed in the course, regardless of the student's high school GPA or age group. The misclassification error rate for the final tree is $0.1606 = 300 / 1,868$.



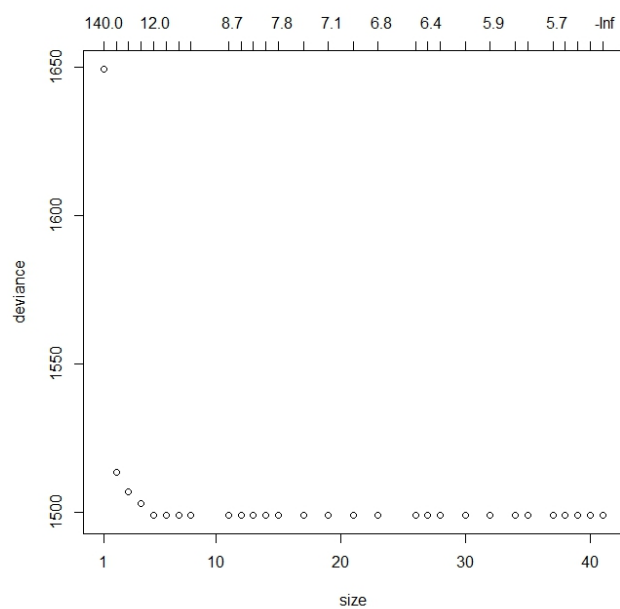
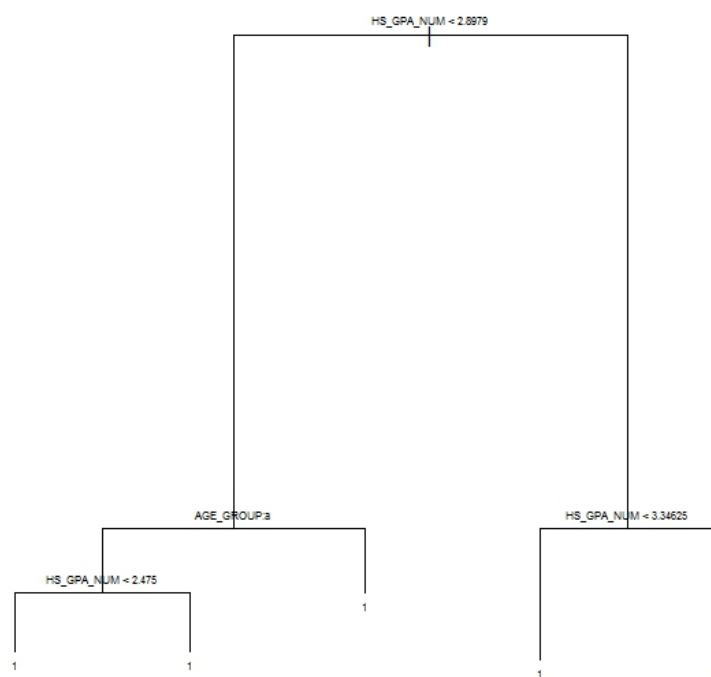


Figure 4: Tree size (horizontal axis) versus cross-validated deviance (vertical axis) for successive prunings of the initial classification tree for students who took English 111 via traditional face-to-face meetings. (The upper scale on the horizontal axis refers to the “cost/complexity” penalty.)



5. Discussion

This study utilized two statistical methods, logistic regression and classification and regression tree (CART) analysis, in an effort to develop predictive models for course success in English 111 based on whether the course was delivered via eLearning or traditional face-to-face meetings. The CART analysis was less successful than logistic regression in producing viable predictive models for the data.

Regarding course success for traditionally delivered English 111, logistic regression produced vastly different results than the CART analysis. The final model that resulted from the logistic regression included more variables than did the final classification tree. Both analyses retained high school GPA and age group in the final models, and while those were the only two variables retained in final classification tree, the final model from the logistic regression also included origin at entry and full-time/part-time status. Because of the viability concern mentioned above, all inferences in the remainder of the discussion will be drawn from the logistic regression analyses.

Evidence suggests that first-time freshmen who took English 111 via traditional face-to-face meetings were 2.3 times more likely to succeed than first-time freshmen who took the course via e-Learning. A future study might include all students who took the course to determine whether this pattern arises across all class standings (Table 4). However, English 111 is a first-year composition course, meant to introduce students to academic conventions in research writing and Standard Written English in preparation for the remainder of their academic career. It is possible that few students wait past their freshmen year to take this course. Students who take the course as freshmen but not in their first semester as college students may be better prepared to handle the responsibility of taking the course via e-Learning. Students who take e-Learning courses must be skilled at establishing their academic routine, managing their time, exhibit self-discipline, and being able to work independently. Perhaps students should be discouraged to take English 111 via e-Learning as

first-time freshmen. If future statistical analyses indicate that students across all class standings who take the course via e-Learning are significantly less likely to succeed in the course than traditional students, it might be sensible to reevaluate the methods of instruction and the tools used to deliver this course via e-Learning in order to assess effectiveness.

High school GPA was the only explanatory variable that was significantly associated with student success in English 111 delivered by both e-Learning and traditional face-to-face meetings. High school GPA is a measure of college preparedness and as would be expected, first-time freshmen with greater high school GPAs are more likely to succeed, which is in agreement with earlier research done by Morris et al. (2005) and Dupin-Bryan (2004).

The only other variable deemed significantly related to course success for students who took English 111 via e-Learning was financial aid status. Students who received financial aid were more likely to succeed in the course than students who did not receive financial aid, which was also in agreement with Morris et al. (2005). At the University of Alaska, financial aid ranges from needs-based to performance-based grants, scholarships, and loans. A minority (38 percent) of the first-time freshmen who took the course via e-Learning received financial aid (Table 1). e-Learning courses usually have a web-based component, requiring students to have access to a computer and a reliable internet connection. These resources can be a financial burden, and students who receive financial aid may be better prepared to handle that burden. Also, internet disruptiveness in Alaska's rural areas might have a negative affect on student motivation and ability to succeed. Further investigation should be done into the types of financial aid students received and its relationship to academic success.

Regarding traditionally delivered English 111, age group and origin at entry were the only two demographic variables that were significant in determining course success. Generally, the probability

of success increased as student's age increased. Evidence did not suggest that out of state students preformed significantly better than Alaskan student. However, relatively fewer first-time freshmen were from out of state than from Alaska (17 versus 1,713; Table 1). The low out of state count may have made it difficult to identify significant differences in course success, when compared to Alaskan first-time freshmen. Foreigner first-time freshmen, on the other hand, performed significantly better than Alaskan students, being 1.5 times more likely to succeed in traditionally delivered English 111 than Alaskan students.

Full-time/part-time status was the only academic variable in addition to high school GPA that was significantly related to course success for traditionally delivered English 111. Students who were enrolled full-time were more likely to succeed in the course than students who were enrolled part-time. It is likely that students who are enrolled part-time have a greater amount of non-academic responsibilities, such as work and family obligations, which negatively impact their academic performance. Since age distribution is showing that students aged 40 or older are performing the best (Table 10), one might speculate that older full-time students produce the best results in traditionally delivered English 111.

One limitation of this study was that there were much fewer first-time freshmen who enrolled in English 111 via e-Learning than those who took the course via traditional face-to-face meetings. It is possible that the lack of reduction in cross-validated deviance in response to increasing tree size for the tree related to course success in English 111 via e-Learning was related to the relatively small sample size for those students.

Another oversight of this study is that it did not consider where students were enrolled in English 111. Each of University of Alaska's major administrative units (MAU), UA Anchorage, UA Fairbanks, and UA Southeast, has its own guidelines for student placement in English 111. Table 13

reports the minimum test scores that each MAU requires students to meet in order to be placed in English 111.

Table 11: Minimum Test Score Required for Placement into English 111 by Major Administrative Unit (MAU) and Placement Test

MAU	Accuplacer			SAT	
	Reading Comprehension	Sentence Skills	Combined	Critical Reading	ACT
UA Anchorage	80	90	170	530	22
UA Fairbanks	85	80	165	510	18
UA Southeast	85	95	180	N/A	N/A

Based on the placement criteria, students at one MAU may be better prepared for the course and thus, more likely to succeed. Moreover, instructional technologies and methodologies may differ among the MAUs.

References

- Blasius, J. & Greenacre, M. (1998). *Visualization of Categorical Data*, San Diego, California: Academic Press.
- Buja, A. & Lee, Y. (2001). Data mining criteria for tree-based regression and classification. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discover and Data Mining*, 27-36. doi: 10.1145/502512.502522
- Davidson, P. (2009). University of Alaska unit cost analysis and other selected issues: Part 3, Distance Education. (Report Digest #45-30033c-09). Alaska State Legislature, Division of Legislative Audit.
- Dupin-Bryant, P. (2004) Pre-entry variables related to retention in online distance education. *American Journal of Distance Education*, 18(4), 199-206.
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Model*. New York, New York: McGraw-Hill/Irwin.
- Lee, Y. & Choi, J. (2011). A review of online course dropout research: implications for practice and future research. *Educational Technology and Research Development*, 59, 593-618. doi: 10.1007/s11423-010-9177-y
- Morris, L. V., Wu, S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *American Journal of Distance Education*. 11(2), 39-54.
- National Center for Education Statistics. (2010). Web tables: Profile of Undergraduate Students: 2007-08. (NCES 2010-205). Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010205>

- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Radford, A. (2011). Learning at a distance: Undergraduate enrollment in distance education courses and degree programs. (NCES 2012-154). National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education. Washington, DC.
- Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge, Great Britain: Cambridge University Press.
- Ripley, B. (2012). tree: Classification and regression trees. R package version 1.0-33. <http://CRAN.R-project.org/package=tree>
- SAS Institute Inc. (1999).
- SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.
- SAS OnlineDoc®, Version 8, Cary, NC. Retrieved from <http://v8doc.sas.com/sashtml/stat/chap3/sect11.htm>
- SAS software, Version 9 of the SAS System for Linux. Copyright © 2001-2010 SAS Institute Inc.
- University of Alaska. (2011). Distance education enhancements banner scheduling manual: How to schedule all your classes with the new distance education parameters. Retrieved from <http://www.alaska.edu/files/studentservices/UA%20DE%20scheduler%20manual.pdf>
- University of Alaska. (2012). Report on University of Alaska e-Learning. (Reference 2). Board of Regents Meeting, Academic and Student Affairs Committee. Retrieved from http://www.alaska.edu/files/bor/121206Ref02_e-Learning_Rpt.pdf
- Xi, D. & Jaggars, S. (2013). Adaptability to online learning: Differences across types of students and academic subject areas. Community College Research Center. New York, NY.

Appendix A

Below are the details for the classification tree depicted in Figure 1, which pertains to course success in English 111 via e-Learning:

node), split, n, deviance, yval, (yprob)

* denotes terminal node

- 1) root 180 221.600 1 (0.3056 0.6944)
 - 2) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,03 Education,04 Engineering,06 Health,07 Letters, Comm., and Philosophy,10 Social Sciences,11 Visual and Performing Arts,12 Vocational Education 165 210.000 1 (0.3333 0.6667)
 - 4) HS_GPA_NUM < 3.094 87 119.200 1 (0.4368 0.5632)
 - 8) DISCIPLINE: 03 Education,04 Engineering,10 Social Sciences,11 Visual and Performing Arts 15 15.010 0 (0.8000 0.2000)
 - 16) HS_GPA_NUM < 2.6509 7 9.561 0 (0.5714 0.4286) *
 - 17) HS_GPA_NUM > 2.6509 8 0.000 0 (1.0000 0.0000) *
 - 9) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,06 Health,07 Letters, Comm., and Philosophy,12 Vocational Education 72 94.180 1 (0.3611 0.6389)
 - 18) SCH_LOAD: 31,51 6 5.407 0 (0.8333 0.1667) *
 - 19) SCH_LOAD: 11,21,41,62,72 66 82.560 1 (0.3182 0.6818)
 - 38) HS_GPA_NUM < 2.535 32 43.860 1 (0.4375 0.5625)
 - 76) TERM: 201201,201203 17 22.070 0 (0.6471 0.3529)
 - 152) AGE_GROUP: 01 Under 20,04 30-39,05 Over 40 5 0.000 0 (1.0000 0.0000) *
 - 153) AGE_GROUP: 02 20-24,03 25-29 12 16.640 0 (0.5000 0.5000)
 - 306) DISCIPLINE: 01 Business and Public Administration,06 Health,12 Vocational Education 7 5.742 0 (0.8571 0.1429) *
 - 307) DISCIPLINE: 02 Computer and Information Science,07 Letters, Comm., and Philosophy 5 0.000 1 (0.0000 1.0000) *
 - 77) TERM: 201103,201202 15 15.010 1 (0.2000 0.8000)
 - 154) RACE: 01 Nat/Ind,06 Other 5 6.730 0 (0.6000 0.4000) *
 - 155) RACE: 03 Asian,05 White 10 0.000 1 (0.0000 1.0000) *
 - 39) HS_GPA_NUM > 2.535 34 34.570 1 (0.2059 0.7941)
 - 78) SCH_LOAD: 11,62 21 26.730 1 (0.3333 0.6667)
 - 156) AGE_GROUP: 01 Under 20,03 25-29,05 Over 40 14 19.120 1 (0.4286 0.5714)
 - 312) HS_GPA_NUM < 2.836 5 6.730 0 (0.6000 0.4000) *
 - 313) HS_GPA_NUM > 2.836 9 11.460 1 (0.3333 0.6667) *
 - 157) AGE_GROUP: 02 20-24,04 30-39 7 5.742 1 (0.1429 0.8571) *
 - 79) SCH_LOAD: 21,41,72 13 0.000 1 (0.0000 1.0000) *
 - 5) HS_GPA_NUM > 3.094 78 81.790 1 (0.2179 0.7821)
 - 10) DISCIPLINE: 01 Business and Public Administration,07 Letters, Comm., and Philosophy,10 Social Sciences,12 Vocational Education 57 69.470 1 (0.2982 0.7018)
 - 20) HS_GPA_NUM < 3.77 50 64.100 1 (0.3400 0.6600)
 - 40) TERM: 201202,201203 21 29.060 0 (0.5238 0.4762)
 - 80) GENDER: Female 12 15.280 0 (0.6667 0.3333)
 - 160) DISCIPLINE: 07 Letters, Comm., and Philosophy 6 5.407 0 (0.8333 0.1667) *
 - 161) DISCIPLINE: 01 Business and Public Administration,10 Social Sciences 6 8.318 1 (0.5000 0.5000) *
 - 81) GENDER: Male 9 11.460 1 (0.3333 0.6667) *
 - 41) TERM: 201103,201201 29 29.570 1 (0.2069 0.7931)
 - 82) TERM: 201103 24 26.990 1 (0.2500 0.7500)
 - 164) RACE: 01 Nat/Ind,03 Asian,06 Other 7 9.561 1 (0.4286 0.5714) *

165) RACE: 05 White 17 15.840 1 (0.1765 0.8235)
330) DEG_LEVEL: 3_Associate 7 0.000 1 (0.0000 1.0000) *
331) DEG_LEVEL: 5_Bachelor 10 12.220 1 (0.3000 0.7000) *
83) TERM: 201201 5 0.000 1 (0.0000 1.0000) *
21) HS_GPA_NUM > 3.77 7 0.000 1 (0.0000 1.0000) *
11) DISCIPLINE: 02 Computer and Information Science,03 Education,04 Engineering,06 Health 21 0.000 1 (0.0000 1.0000) *
3) DISCIPLINE: 05 Foreign Languages,08 Math, Physical and Life Sciences,14 Interdisciplinary Studies 15 0.000 1 (0.0000 1.0000) *

Appendix B

Below are the details for the classification tree depicted in Figure 3, which pertains to course success in English 111 via traditional face-to-face meetings:

node), split, n, deviance, yval, (yprob)

* denotes terminal node

```

1) root 1868 1646.000 1 ( 0.160600 0.839400 )
2) HS_GPA_NUM < 2.8979 596 738.400 1 ( 0.310403 0.689597 )
4) AGE_GROUP: 01 Under 20 415 543.000 1 ( 0.361446 0.638554 )
8) HS_GPA_NUM < 2.475 157 217.500 1 ( 0.484076 0.515924 )
16) SCH_LOAD: 21,41,51,62,72 151 209.300 0 ( 0.503311 0.496689 )
32) TERM: 201103,201201 72 96.230 0 ( 0.611111 0.388889 )
64) DISCIPLINE: 03 Education,07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,09
Natural Resources,10 Social Sciences 41 47.690 0 ( 0.731707 0.268293 ) *
65) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,04
Engineering,06 Health,11 Visual and Performing Arts,12 Vocational Education,14 Interdisciplinary Studies 31 42.680
1 ( 0.451613 0.548387 ) *
33) TERM: 201203 79 106.700 1 ( 0.405063 0.594937 ) *
17) SCH_LOAD: 11,31,82 6 0.000 1 ( 0.000000 1.000000 ) *
9) HS_GPA_NUM > 2.475 258 309.200 1 ( 0.286822 0.713178 )
18) SCH_LOAD: 21,51,82 26 35.890 0 ( 0.538462 0.461538 )
36) HS_GPA_NUM < 2.6059 8 6.028 0 ( 0.875000 0.125000 ) *
37) HS_GPA_NUM > 2.6059 18 24.060 1 ( 0.388889 0.611111 )
74) DISCIPLINE: 04 Engineering,07 Letters, Comm., and Philosophy,09 Natural Resources,11 Visual and
Performing Arts 13 17.940 0 ( 0.538462 0.461538 ) *
75) DISCIPLINE: 03 Education,06 Health,08 Math, Physical and Life Sciences,12 Vocational Education 5
0.000 1 ( 0.000000 1.000000 ) *
19) SCH_LOAD: 11,31,41,62,72 232 265.200 1 ( 0.258621 0.741379 )
38) HS_GPA_NUM < 2.881 225 250.300 1 ( 0.244444 0.755556 )
76) DISCIPLINE: 02 Computer and Information Science,05 Foreign Languages,10 Social Sciences 22 30.500
1 ( 0.500000 0.500000 ) *
77) DISCIPLINE: 01 Business and Public Administration,03 Education,04 Engineering,06 Health,07 Letters,
Comm., and Philosophy,08 Math, Physical and Life Sciences,09 Natural Resources,11 Visual and Performing Arts,12
Vocational Education,14 Interdisciplinary Studies 203 212.200 1 ( 0.216749 0.783251 )
154) HS_GPA_NUM < 2.7635 120 141.200 1 ( 0.275000 0.725000 )
308) HS_GPA_NUM < 2.7079 98 104.400 1 ( 0.224490 0.775510 ) *
309) HS_GPA_NUM > 2.7079 22 30.500 1 ( 0.500000 0.500000 )
618) DISCIPLINE: 03 Education,04 Engineering,07 Letters, Comm., and Philosophy,08 Math, Physical and
Life Sciences,14 Interdisciplinary Studies 15 19.100 0 ( 0.666667 0.333333 ) *
619) DISCIPLINE: 01 Business and Public Administration,06 Health,11 Visual and Performing Arts 7
5.742 1 ( 0.142857 0.857143 ) *
155) HS_GPA_NUM > 2.7635 83 64.930 1 ( 0.132530 0.867470 )
310) HS_GPA_NUM < 2.79085 18 0.000 1 ( 0.000000 1.000000 ) *
311) HS_GPA_NUM > 2.79085 65 59.110 1 ( 0.169231 0.830769 )
622) DISCIPLINE: 01 Business and Public Administration,07 Letters, Comm., and Philosophy,08 Math,
Physical and Life Sciences,11 Visual and Performing Arts,12 Vocational Education 51 53.180 1 ( 0.215686 0.784314 )
*
623) DISCIPLINE: 03 Education,04 Engineering,06 Health,09 Natural Resources 14 0.000 1 ( 0.000000
1.000000 ) *
```

- 39) HS_GPA_NUM > 2.881 7 8.376 0 (0.714286 0.285714) *
- 5) AGE_GROUP: 02 20-24,03 25-29,04 30-39,05 Over 40 181 177.800 1 (0.193370 0.806630)
- 10) AGE_GROUP: 02 20-24,03 25-29,05 Over 40 162 169.100 1 (0.216049 0.783951)
- 20) SCH_LOAD: 11,31,41,51,62,82 118 135.900 1 (0.262712 0.737288)
- 40) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,06 Health,07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,10 Social Sciences,11 Visual and Performing Arts,12 Vocational Education,14 Interdisciplinary Studies 100 122.200 1 (0.300000 0.700000)
- 80) GENDER: Female 52 50.910 1 (0.192308 0.807692)
- 160) RACE: 01 Nat/Ind,05 White 36 42.540 1 (0.277778 0.722222) *
- 161) RACE: 03 Asian,04 Black,06 Other 16 0.000 1 (0.000000 1.000000) *
- 81) GENDER: Male 48 65.200 1 (0.416667 0.583333) *
- 41) DISCIPLINE: 03 Education,04 Engineering,09 Natural Resources 18 7.724 1 (0.055556 0.944444) *
- 21) SCH_LOAD: 21,72 44 26.810 1 (0.090909 0.909091)
- 42) DISCIPLINE: 07 Letters, Comm., and Philosophy,11 Visual and Performing Arts,12 Vocational Education 18 19.070 1 (0.222222 0.777778) *
- 43) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,04 Engineering,05 Foreign Languages,06 Health,08 Math, Physical and Life Sciences,10 Social Sciences,14 Interdisciplinary Studies 26 0.000 1 (0.000000 1.000000) *
- 11) AGE_GROUP: 04 30-39 19 0.000 1 (0.000000 1.000000) *
- 3) HS_GPA_NUM > 2.8979 1272 772.100 1 (0.090409 0.909591)
- 6) HS_GPA_NUM < 3.34625 504 427.500 1 (0.150794 0.849206)
- 12) DISCIPLINE: 05 Foreign Languages,12 Vocational Education 30 40.380 1 (0.400000 0.600000)
- 24) GENDER: Female 13 16.050 0 (0.692308 0.307692) *
- 25) GENDER: Male 17 15.840 1 (0.176471 0.823529) *
- 13) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,03 Education,04 Engineering,06 Health,07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,09 Natural Resources,10 Social Sciences,11 Visual and Performing Arts,14 Interdisciplinary Studies 474 375.200 1 (0.135021 0.864979)
- 26) SCH_LOAD: 11,21,51 27 35.590 1 (0.370370 0.629630)
- 52) DISCIPLINE: 07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,11 Visual and Performing Arts 11 14.420 0 (0.636364 0.363636) *
- 53) DISCIPLINE: 01 Business and Public Administration,04 Engineering,06 Health,10 Social Sciences 16 15.440 1 (0.187500 0.812500)
- 106) HS_GPA_NUM < 3.1318 9 0.000 1 (0.000000 1.000000) *
- 107) HS_GPA_NUM > 3.1318 7 9.561 1 (0.428571 0.571429) *
- 27) SCH_LOAD: 31,41,62,72,82 447 329.500 1 (0.120805 0.879195)
- 54) RACE: 01 Nat/Ind,02 Hi/Pac Is,04 Black,06 Other 185 164.000 1 (0.162162 0.837838)
- 108) DISCIPLINE: 01 Business and Public Administration,03 Education,04 Engineering 57 71.100 1 (0.315789 0.684211)
- 216) FINAID_FLAG: 0 26 18.600 1 (0.115385 0.884615) *
- 217) FINAID_FLAG: 1 31 42.940 1 (0.483871 0.516129) *
- 109) DISCIPLINE: 02 Computer and Information Science,06 Health,07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,09 Natural Resources,10 Social Sciences,11 Visual and Performing Arts,14 Interdisciplinary Studies 128 79.650 1 (0.093750 0.906250) *
- 55) RACE: 03 Asian,05 White 262 160.500 1 (0.091603 0.908397)
- 110) DISCIPLINE: 02 Computer and Information Science,07 Letters, Comm., and Philosophy,08 Math, Physical and Life Sciences,10 Social Sciences,11 Visual and Performing Arts 153 114.800 1 (0.124183 0.875817) *
- 111) DISCIPLINE: 01 Business and Public Administration,03 Education,04 Engineering,06 Health,09 Natural Resources,14 Interdisciplinary Studies 109 40.590 1 (0.045872 0.954128) *
- 7) HS_GPA_NUM > 3.34625 768 308.400 1 (0.050781 0.949219)
- 14) DISCIPLINE: 02 Computer and Information Science,05 Foreign Languages,06 Health,07 Letters, Comm., and Philosophy,10 Social Sciences,11 Visual and Performing Arts,14 Interdisciplinary Studies 379 204.800 1 (0.076517 0.923483)
- 28) RACE: 01 Nat/Ind,03 Asian 67 59.840 1 (0.164179 0.835821)

56) DISCIPLINE: 02 Computer and Information Science,07 Letters, Comm., and Philosophy,10 Social Sciences
 48 51.670 1 (0.229167 0.770833)
 112) HS_GPA_NUM < 3.7525 31 38.990 1 (0.322581 0.677419) *
 113) HS_GPA_NUM > 3.7525 17 7.606 1 (0.058824 0.941176) *
 57) DISCIPLINE: 05 Foreign Languages,06 Health,11 Visual and Performing Arts 19 0.000 1 (0.000000
 1.000000) *
 29) RACE: 02 Hi/Pac Is,04 Black,05 White,06 Other 312 137.600 1 (0.057692 0.942308)
 58) SCH_LOAD: 11,72 101 69.530 1 (0.108911 0.891089) *
 59) SCH_LOAD: 21,31,41,51,62,82 211 61.450 1 (0.033175 0.966825) *
 15) DISCIPLINE: 01 Business and Public Administration,03 Education,04 Engineering,08 Math, Physical and Life
 Sciences,09 Natural Resources,12 Vocational Education 389 92.960 1 (0.025707 0.974293)
 30) RACE: 01 Nat/Ind,02 Hi/Pac Is,03 Asian,06 Other 154 62.900 1 (0.051948 0.948052) *
 31) RACE: 04 Black,05 White 235 23.050 1 (0.008511 0.991489)
 62) HS_GPA_NUM < 3.425 29 14.560 1 (0.068966 0.931034)
 124) HS_GPA_NUM < 3.4095 21 0.000 1 (0.000000 1.000000) *
 125) HS_GPA_NUM > 3.4095 8 8.997 1 (0.250000 0.750000) *
 63) HS_GPA_NUM > 3.425 206 0.000 1 (0.000000 1.000000) *

Appendix C

Below are the details for the classification tree depicted in Figure 5, which pertains to course success in English 111 via traditional face-to-face meetings:

node), split, n, deviance, yval, (yprob)

* denotes terminal node

- 1) root 1868 1646.0 1 (0.16060 0.83940)
- 2) HS_GPA_NUM < 2.8979 596 738.4 1 (0.31040 0.68960)
- 4) AGE_GROUP: 01 Under 20 415 543.0 1 (0.36145 0.63855)
 - 8) HS_GPA_NUM < 2.475 157 217.5 1 (0.48408 0.51592) *
 - 9) HS_GPA_NUM > 2.475 258 309.2 1 (0.28682 0.71318) *
- 5) AGE_GROUP: 02 20-24,03 25-29,04 30-39,05 Over 40 181 177.8 1 (0.19337 0.80663) *
- 3) HS_GPA_NUM > 2.8979 1272 772.1 1 (0.09041 0.90959)
- 6) HS_GPA_NUM < 3.34625 504 427.5 1 (0.15079 0.84921) *
- 7) HS_GPA_NUM > 3.34625 768 308.4 1 (0.05078 0.94922) *