Journal of Undergraduate Research at Minnesota State University, Mankato

Volume 14 Article 6

2014

Modeling Student Engagement in the Classroom

Sarah Painter
Minnesota State University, Mankato

Modeling Student Engagement in the Classroom

Sarah Painter (Department of Mathematics)

Dr. Rebecca Bates (Department of Integrated Engineering)

Abstract

Connections to Community is a multi-institutional study that looks at the influence of community on post-secondary, science and engineering students and their engagement in academic activity. This paper focuses specifically on student engagement within the classroom as a follow-up to a previous paper by Wendy Hoffman, *Identifying Influential Variables of Student Academic Engagement* (Hoffman, 2013). The goal of this work is to model student engagement in the classroom using classroom observation data that has been cleaned and then compare the results with those found in Hoffman's paper which used pre-cleaning data. The cleaned data is used to create two data sets, one with any observations with missing values removed completely and one with the use of the variable median to replace missing values. These sets are then used to create 12 predictor constructs and one response construct that are thought to be valid according to education experts. Two starting models are created from each data set for a total of four starting models. A form of backward elimination linear regression is used on the four starting models to create four reduced models. The residuals of each reduced model are tested using the D'Agostino-Pearson test statistic, and then the models are compared using the PRESS and SSE statistics. We find that *Students' Reluctance to Leave* and *Students' Discussions with Peers* are influential in determining student engagement within the classroom. Also, the level students are at within their program and whether students are automotive engineering technology or chemistry students is important. The results found here generally coincide with the findings from the previous paper, but there are differences in the data set that uses the median to replace missing values. Where Hoffman found reliable results for models using this data set, this paper finds results less credible. More work is needed to understand what happened during the cleaning process to create these differences.

1. Introduction and Background

Researchers in engineering education have been examining the issues of student engagement, persistence, and retention within science, technology, engineering, and mathematics (STEM) for many years. It seems that although STEM and non-STEM undergraduates alike tend to move between majors, it's unlikely that non-STEM students switch into STEM programs, leaving the STEM programs with ultimately fewer students while other disciplines may grow in population (Seymour, 2002). This has led experts in engineering education to look into persistence and retention. What is it that causes this net loss in engineering majors? What can be done to prevent it? While one way to address this problem is to increase the ease of moving into STEM majors, this paper focuses on the student experience within STEM majors. It is believed by experts that persistence and retention are highly correlated with student engagement, which is the focus of this paper.

1.1 Engagement and the Classroom

Student engagement is an integral piece of the education process. It is believed to be directly linked to student achievement (Kim, et al., 2012). Students' connections to community are also thought to be influential in determining student engagement. The strength of connections to community is impacted by three main areas: academic/cognitive orientation, sense of support, and the classroom (Floyd-Smith, et al., 2010). The classroom experience is indicated by course format, classroom characteristics, course content, instructor efforts to connect, effective use of questions, and instructor preparation (Floyd-Smith, et al., 2010).

The "Connections to Community" (CTC) project looks to find the relationship between community and engagement and its role in the success of STEM students. It is a five year study involving five universities across the United States: a large research-based university in the Pacific Northwest, a medium teaching-based university in the Midwest, a historically black university in the Southeast, a women's college in the Northeast, and a faith-based institution in the Pacific Northwest. Data collection has involved surveys, focus groups, interviews, and classroom observations. This paper uses just a small piece of the data collected for the study. The work presented here is based on classroom observation data collected between August 2010 and May 2012. Other work from this study was used as reference for this paper (Allendoerfer, et al., 2012; Bates, et al., 2013; Floyd-Smith, et al., 2010; Floyd-Smith, et al., 2012; Kim, et al., 2012; Plett, et al., 2011; Wilson, et al., 2013).

Finding a model for student engagement using classroom observation data allows one to see what does or does not happen in the classroom that is influential in determining the engagement of students during learning. If we can know more about what helps students be more engaged in their classroom learning experiences, we can make an effort to do those things more, possibly increasing the number of STEM undergraduate student successes.

1.2 Previous Paper

Throughout this paper, Wendy Hoffman's paper, *Identifying Influential Variables of Student Academic Engagement*, (Hoffman, 2013) will be referred to repeatedly. This previous paper was also written in conjunction with the larger CTC study and looked to model student engagement within the classroom. Hoffman had several questions:

- What is the best method to model Likert scaled data, specifically measurable classroom engagement characteristics?
- What controls are most influential when predicting student engagement?
- What characteristics of the classroom are most influential when predicting student engagement?

Also hypothesized was that the way missing values were handled would significantly affect the model. This led to two methods of dealing with missing values that created the need for two different data sets. Data Set 1 included only observations without missing values, whereas Data Set 2 replaced any missing values with the median of the respective column variable. In order to test whether control variables are related to student engagement, each data set was used to create a model with control variables included and another model without control variables included. This resulted in four starting models: Base 1, Control 1, Base 2, and Control 2. Base 1 used input from Data Set 1 and included only base variables (non-controls observed directly within the classroom). Control 1 used input from Data Set 1 and included both the base and control variables. Base 2 used input from Data Set 2 and included only base variables. Finally, Control 2 used input from Data Set 2 and included both base and control variables.

After trying several methods of regression analysis, Hoffman found that linear regression was the best fit for the data. Backward regression techniques were used to reduce the four models. For each of the resulting final four models, the normality of the residuals was tested using four test statistics as well as QQ-plots. This was needed in order to verify the assumptions of linear regression. The four test statistics included two versions of the D'Agostino-Pearson test statistic (one using the Chi-square distribution with two degrees of freedom and one using the empirical distribution) and two versions of the Jarque-Bera test statistic (one using the Chi-square distribution with two degrees of freedom and one using the empirical distribution) (Rahman & Wu, 2013). After verifying the normality of the residuals, the reduced models were then compared using a combination of the predicted residual sum of squares (PRESS) and the error sum of squares (SSE) statistics (Dielman, 1996), and it was found that the final Base 1 and Control 2 models were the best for modeling student engagement in the classroom. The final Base 1 model included the constructs: *Questions Asked by Students*, *Students' Reluctance to Leave*, *Students' Discussions with Peers*, and *Amount of Theoretical Focus of Each Observation*. The final Control 2 model included the control variable *Level Code* which represents year in program and the constructs: *Questions Asked by Students* and *Students' Discussions with Peers*. The constructs will be explained fully in Section 2.1 (The Data). Since both models used *Questions Asked by Students* and *Students' Discussions with Peers*, it seems that these constructs are important in predicting student engagement within the classroom. Also, both of the reduced Control models Hoffman found included *Level Code*, meaning a student's program level may be the most influential control variable for predicting student engagement.

After Hoffman's original analysis, the classroom observation data underwent a thorough cleaning process which will be discussed further in Section 2.1. The newly cleaned data was then used for analysis in this paper. Here, Hoffman's results will be compared to results found using this "clean" data and similar methods.

1.3 Research Questions

The central question this paper addresses is: *are there differences between the pre-cleaning data and the post-cleaning data?* Using this newly cleaned data, questions similar to Hoffman's (Hoffman, 2013) are asked:

Question 1: What control variables (if any) most influence student engagement?

Question 2: What classroom characteristics most influence student engagement?

Question 3: Does the treatment of missing values affect the models?

These questions will be addressed using methods similar to Hoffman's. This will allow a direct comparison between results found in each paper.

1.4 Paper Organization

In the next section, the methodology will be covered. First, the data will be described in detail: the collection process, the types of population sampling, the measuring instrument, the cleaning process, and the defining of constructs and sets. The theory behind analytical methods will then be covered: the method of linear regression analysis, model reduction methods, normality testing methods, and model comparison methods. Following methodology, the results of analysis will be covered. These results will then be discussed within the context of our research questions. Final conclusions and possible future directions will be discussed in Section 5.

2. Methodology

This section provides an in-depth look at the data used for this paper, as well as the theory behind the methods used for analysis of the data.

2.1 The Data

The data used for this portion of the study consists of classroom observations taken between August 2010 and May 2012 from only two universities involved in the study: school A being a teaching-based institution in the Midwest and school B being a research-based institution in the Pacific Northwest. Observations were taken within Automotive, Chemical, Civil, Electrical, and Mechanical Engineering classes, as well as related courses in Math, Chemistry, Computer Science, and Physics. The classes were also classified according to level (sophomore, junior, or senior) within their respective programs. Because the classroom environment may change throughout the semester, most classes were observed three times.

The observation instrument was created by the CTC researchers, using material from the University of Minnesota's Center for Teaching and Learning (Classroom Observation Instruments). The survey was divided into three sections: student behaviors (E), characteristics of instruction (I), and classroom characteristics (C). Within each section, the questions were answered using a Likert scale from 0 to 4 or 1 to 4. The classroom observation form included a total of 40 base variables (those directly observed within the classroom) and 7 control variables (faculty, institution, date, format, etc.) When the observation forms were originally entered into the computer, it was decided that several other control variables should be added to the description of each observation. These included the institution, the teaching style observed, major/program for which the class was required, and the observer. The possible observed teaching styles included *sage on the stage*, *connected sage on the stage*, and *sage off the stage*. *Sage on the stage* refers to the traditional style of teaching where the professor stands up front and lectures. *Connected sage on the stage* refers to a traditional lecture style but with more interaction between the professor and students. Here, the instructor may involve the students in the process by asking the students questions or doing problems together. *Sage off the stage* refers to the more unconventional style of teaching where the professor steps down from the podium and allows the students to take the lead in the learning process (also referred to as *Guide on the side*).

The original data used in Hoffman's paper underwent a rigorous cleaning process before being used again for this paper. This process included checking the entered data with the original observation forms twice, gathering outside information to fill in missing values, and clearing up ambiguities in the data. For example, there were some inconsistencies in the wording of the variable E3. In some forms, the question was worded: "Do students express body language indicative of being bored or disinterested?", but in others, the question read: "Do students express body language indicative of being engaged or interested?" Therefore, the wording of E3 was recorded for each observation and then certain observations were reverse coded to make the variable fully consistent. Questions E8 and E9 dealt with students' participation in group activities, but group activity was rarely present during observations making E8 and E9 irrelevant for any use in analysis. Problems such as these were fixed during the cleaning process to create a more valid data set with which to work. Again, the newly cleaned data was used in the analysis for this paper. See Appendix A for the full list of variables and their associated questions.

The 40 base variables were reduced down to 12 constructs and 1 response variable according to the expertise of educational theory. Both attentiveness and attendance have been shown to be good measurements of engagement (Kim, et al., 2012). Here, as in the previous paper, engagement is measured in terms of attentiveness. Informed by the theory, the variables E1, E2, E3, and E4 were averaged to create the response variable *Attentiveness*. The other 12 constructs used as predictors in the models were also created by averaging groups of variables. Again, these groupings were chosen based on theory. These groupings are similar to the groupings used in Hoffman's paper (Hoffman, 2013).

Each construct was tested for reliability using Cronbach's coefficient α. The α-values give a measure for internal reliability within each construct. A higher α-value implies better consistency within the variables grouped to create the construct. A value of 0.7 or higher is generally considered acceptable for use with Likert-scaled data (Gliem & Gliem, 2003) . The α-values were computed in SPSS

DOI: 10.56816/2378-6949.1050

(http://www-01.ibm.com/software/analytics/spss/products/statistics/index.html) for each of the constructs used in analysis.

The constructs used in analysis are given in Table 1, as well as their α-values and variables used for creation. Certain constructs include only one variable and therefore need not be tested for internal reliability. Table 2 gives the control variables used in analysis.

Table 1. Constructs with respective grouping variables and α-values for Data Sets 1 and 2. Data Set 1 consists of all observations without any missing values. Data Set 2 consists of all observations, replacing any missing values with their variable median.

Construct	**Variables**	**α_1 (Data Set 1)**	**α_2 (Data Set 2)**
Questions Asked by Students	E6, E7	0.663	0.588
Students' Reluctance to Leave	E11	---	---
Students' Discussions with Peers	E12, E13	0.812	0.741
Instructor Availability	I18	---	---
Physical Environment of the Classroom	C4	---	---
Amount of Theoretical Focus	C5	---	---
Amount of Application Focus	C6	---	---
Amount of Design Focus	C7	---	---
Questions Posed by Instructor	I6, I7, I20	0.663	0.638
Instructor Efforts to Connect with Students	I2, I3, I9	0.547	0.551
Instructional Delivery	I1, I4, I10, I11, I12, I13, I14, I15, I16, I17, I19	0.826	0.813
Supporting Documents	C1, C2, C3	0.948	0.933
Response	E1, E2, E3, E4	0.792	0.760

Table 2. Control variables with respective coding/labeling schemes

Variable	**Coding/Labeling Scheme**
Institution Code (nominal)	1 = School A, 2 = School B
Teaching Style Code (nominal)	1 = connected sage on stage, 2 = sage off stage, 3 = sage on stage
Automotive Engineering Technology (dichotomous)	0 = not required for program, 1 = required for program
Electrical Engineering (dichotomous)	0 = not required for program, 1 = required for program
Civil Engineering (dichotomous)	0 = not required for program, 1 = required for program
Mechanical Engineering (dichotomous)	0 = not required for program, 1 = required for program
Chemical Engineering (dichotomous)	0 = not required for program, 1 = required for program
Chemistry (dichotomous)	0 = not required for program, 1 = required for program
Program Level (ordinal)	2 = sophomore, 3 = junior, 4 = senior
Observer Code (nominal)	Anonymous, labeled 1, 2, 3, 4, 5 for each observer

The creation of constructs here follows the same method used by Hoffman in the previous paper. This ensures that comparing the results of analysis with Hoffman's results still makes sense. However, the data has been cleaned since Hoffman's analysis, and therefore the constructs created here differ in value from the constructs created by Hoffman.

The definition of data sets and models used by Hoffman (Hoffman, 2013) is also used here. Data Set 1 includes all observations without any missing values. Data Set 2 includes all observations with missing values replaced by the variable median. The Base Models include only the base (directly observed) variables. The Control Models include both control and base variables. This creates four separate starting models: Base Model 1, Base Model 2, Control Model 1, and Control Model 2. Again, this is to test the hypotheses that the treatment of missing values matters and that a significant relationship exists between the control variables and engagement.

With clean data and newly created constructs, linear regression analysis is used to find a model for predicting student engagement in the classroom.

2.2 Linear Regression Analysis

Since the purpose of this paper is to compare the "clean" data with the "dirty" data, methods of analysis similar to Hoffman's are used. Hoffman found linear regression to be the most appropriate form of regression with which to model the data. Therefore, linear regression analysis is used here as well.

Let $\mathbf{X} = (\mathbf{X}_1,\ \mathbf{X}_2, \dots,\ \mathbf{X}_\mathrm{m})$ be the matrix of independent variables within the linear model where $\mathbf{X}_1,\ \mathbf{X}_2, \dots,\ \mathbf{X}_\mathrm{m}$ denote vectors of measurements of each input variable where $j = 0, 1, \dots, m$ is the number of independent variables being considered within the respective model and $\mathbf{X}_\mathrm{j} = (x_{1j},\ x_{2j}, \dots,$

x_{nj}) for $i = 1, 2, \dots, n$ is the number of observations. Let $\mathbf{Y}$ be the vector of response variables where $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$ and each observed $y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_n x_{in} + e_i$ with $e_i = y_i - (\beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_n x_{in})$. However, the true values of β_0, $\beta_1, \dots,$ β_n, and e_i are unknown. Therefore, one computes, $e_i = y_i - \hat{y}_i$, the residuals of the linear model, where $\hat{y}_i = \hat{\beta}_0 x_{i0} + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_n x_{in}$ is the equation of the line that approximates the relationship between $\mathbf{X}$ and $\mathbf{Y}$.

In order for linear regression analysis to be appropriate, the given data must satisfy these four assumptions (Dielman, 2005):

1. Homoscedasticity; common but generally unknown variance, $var(e_i) = \sigma^2$ for $i = 1, 2, \dots, n$
2. Residuals follow a normal distribution
3. Residuals have a mean value of zero; $E(e_i) = 0$ for $i = 1, 2, \dots, n$
4. Errors are mutually uncorrelated; $cov(e_i, e_j) = 0$ for all $i \neq j$

If the residuals do in fact follow a normal distribution, the observed response variables $y_1, y_2, \dots, y_n$ must be independent, follow a normal distribution, and have a common variance (Kutner, Nachtsheim, & Neter, 2004, p. 218). Since the residuals are simply linear combinations of the response variables, the residuals must also be independent with a common variance (Kutner, Nachtsheim, & Neter, 2004, pp. 644-647). The third assumption, $E(e_i) = 0$ for $i = 1, 2, \dots, n$, is automatically satisfied by the method of least squares (Kutner, Nachtsheim, & Neter, 2004, p. 102). Therefore, the residuals computed from the final reduced models need only be tested for normality.

The method of backward elimination is used here to reduce each of the four starting models mentioned earlier (Control 1, Control 2, Base 1, and Base 2). Originally, all of the variables applicable to the respective model are used to create $\mathbf{X}$, the input matrix used to predict $\mathbf{Y}$. The β-coefficient and p-value are computed for each variable included in the input matrix. The β-coefficient signifies the strength of the relationship between the corresponding input variable and student engagement. The input variable with the highest p-value is then removed from the input matrix $\mathbf{X}$. This is just one step in the backward elimination process. This process continues until all of the variables left have a p-value less than or equal to 0.01. This method is performed on each of the four starting models to obtain four reduced models.

The normality of the residuals is tested in order to verify the four assumptions stated earlier. Therefore, the residuals e_i are computed at the last step of backward elimination for each model and then tested for normality. There are many possible methods for testing. Here, the D'Agostino-Pearson approach is used.

2.3 D'Agostino-Pearson Test for Normality

The D'Agostino-Pearson method uses sample estimates $\sqrt{b_1}$ and b_2 to test the normality of a data sample. Let $(X_1, \dots, X_n)$ be a sample of n observations, M_j be the jth sample central moment, and $\bar{X}$ be the sample mean. Then $\sqrt{b_1}$ and b_2 are defined as follows:

$$\sqrt{b_1} = \frac{M_3}{\sqrt{M_2^3}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{3/2}}$$

$$b_2 = \frac{M_4}{M_2^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{2}}$$

where $M_j = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^j$.

There are two forms of the D'Agostino-Pearson statistic: DPC and DPE. DPC uses the Chi-square distribution with two degrees of freedom to obtain percentiles and the upper tail p-value, whereas DPE uses the empirical distribution. In Hoffman's paper (Hoffman, 2013), four test statistics were used: DPC, DPE, and two forms of the Jarque-Bera test statistic. Rahman and Wu (Rahman & Wu, 2013) found DPE to be more reliable than DPC and the two forms of the Jarque-Bera statistic. Therefore, DPE alone is used here to test the normality of the residuals found for each reduced model.

To compute DPE (Rahman & Wu, 2013):

1. Simulate samples of size n (sample size of the data) from $N(0,1)$, compute $\sqrt{b_1}$ and b_2, store the values.
2. Separately, take samples of size n from $N(0,1)$, compute $\sqrt{b_1}$ and b_2, obtain the percentile positions of $\sqrt{b_1}$ and b_2 in the stored respective empirical distributions in Step 1, compute: $\mathrm{DPE} = X^2(\sqrt{b_1}) + X^2(b_2)$ and store, where $X(\sqrt{b_1})$ and $X(b_2)$ are defined as the standard normal score for the respective percentile position for $\sqrt{b_1}$ and b_2, respectively.
3. Compute $\sqrt{b_1}$ and b_2 for the data, obtain percentile positions of $\sqrt{b_1}$ and b_2 in the stored respective empirical distributions in Step 1, compute $\mathrm{DPE} = X^2(\sqrt{b_1}) + X^2(b_2)$ and then obtain the upper tail p-value by using the empirical distribution in Step 2.

The null hypothesis is H_0: the data is normally distributed, whereas the alternative hypothesis is H_1: the data is not normally distributed. Since a higher p-value indicates that one should accept the null hypothesis and a p-value less than or equal 0.05 indicates that one should reject the null hypothesis with 95% confidence, a p-value greater than 0.05 will be sufficient for verifying normality.

DOI: 10.56816/2378-6949.1050

After verifying the normality of the residuals for each reduced model, the models are compared to find the 'best' model. For this, a combination of two statistics is used: the predicted residual sum of squares (PRESS) and the general error sum of squares (SSE).

2.4 PRESS and SSE Statistics for Model Comparison

The predicted residual sum of squares (PRESS) statistic and the error sum of squares (SSE) statistic are both measurements of error for the model. The PRESS statistic considers the difference between the predicted value calculated using the previously found linear equation and the predicted value calculated using an estimation of the linear equation after removing one row of the original data. This assesses the ability of the model to accurately predict future observations. The SSE statistic is simpler in that it considers the difference between the value of an output value from the original data and its respective predicted value using the previously found linear model, assessing the ability of the model to accurately predict know observations.

PRESS is defined as

$$PRESS = \sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-1}\right)^2,$$

where y_i is the ith predicted value using the linear equation from the previously found model, and $\hat{y}_{i,-1}$ is the ith predicted value using an estimation of the model with the ith row removed from the original data (Dielman, 1996, p. 117).

SSE is defined as

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$

where y_i is the ith given output value in the original data, and $\hat{y}_i$ is the ith predicted value using the linear equation from the previously found model (Dielman, 1996, p. 103).

Following Hoffman's approach, both statistics are computed and compared for each of the four reduced models. A model is considered 'better' if the values of its respective PRESS and SSE statistics are closer together (Kutner, Nachtsheim, & Neter, 2004, pp. 373-374). This is the primary method for comparing the four reduced models, but results found at other steps of analysis are also considered.

3. Results

The final reduced models using backward elimination are given in Table 3. The β-coefficients are given as a numeric representation of the relationship between the input variable and engagement, the Z-scores for completeness, and the p-values for their use in the elimination process. Several transformations were used on the response variable of *Attentiveness* in each model in order to obtain normal residuals in the resulting reduced models. The Control 1 model was found to have normal residuals without transforming the response variable. The Base 1 model was found to have normal residuals only after using the transformation $y = \sqrt{x}$ on the response variable. The reduced Control 2 and Base 2 models given in Table 3 were found using the transformation $y = \left(\frac{x}{2}\right)^2$.

Table 3. Input variables and corresponding β-coefficients, Z-scores, and p-values for final models

Control 1			
Variables	**β-coefficient**	**Z-score**	**p-value**
Intercept	2.502	6.039	1.55E-09
Automotive Engineering Technology	-0.487	-2.879	0.004
Chemistry	-0.487	-2.833	0.005
Program Level	-0.213	-2.761	0.006
Students' Reluctance to Leave	0.167	3.436	0.001
Students' Discussions with Peers	-0.220	-2.645	0.008
Instructional Delivery	0.478	4.151	3.31E-05

Control 2			
Variables	**β-coefficient**	**Z-score**	**p-value**
Intercept	1.907	5.129	2.91E-07
Automotive Engineering Technology	-0.741	-4.264	2.01E-05
Students' Reluctance to Leave	0.185	3.736	0.0001
Students' Discussions with Peers	-0.299	-4.089	4.34E-05
Amount of Theoretical Focus	-0.161	-3.301	0.001
Instructor Efforts to Connect with Students	0.478	4.976	6.50E-07

Base 1			
Variables	**β-coefficient**	**Z-score**	**p-value**
Intercept	1.498	14.859	0
Students' Reluctance to Leave	0.048	3.018	0.003
Students' Discussions with Peers	-0.129	-5.163	2.44E-07
Questions Posed by the Instructor	0.1250	4.049	5.14E-05

Base 2			
Variables	**β-coefficient**	**Z-score**	**p-value**
Intercept	1.880	4.879	1.07E-06
Students' Reluctance to	0.186	3.622	0.0003

Leave			
Students' Discussions with Peers	-0.300	-3.960	7.51E-05
Amount of Theoretical Focus	-0.155	-3.064	0.002
Instructor Efforts to Connect with Students	0.447	4.506	6.61E-06

The p-values for the DPE test statistic are given in Table 4. Since all four p-values are greater than 0.05, we choose to accept the null hypothesis and conclude that the residuals follow a normal distribution.

Table 4. Normality of residuals of final models (DPE)

Model	**p-value for DPE**
Control 1	0.6242
Control 2	0.0996
Base 1	0.1981
Base 2	0.0505

The values of the PRESS and SSE statistics as well as their computed differences are given in Table 5. A smaller difference implies a more reliable model.

Table 5. PRESS and SSE values and differences

Model	**PRESS**	**SSE**	**Difference**
Control 1	13.377	36.888	23.511
Control 2	25.190	173.830	148.639
Base 1	4.24E-01	4.045	3.621
Base 2	10.438	187.690	177.251

4. Discussion

The results of analysis give four final models (listed in Table 3), a statistic measuring normality for each of those models (listed in Table 4), and a statistic measuring reliability for each of the models (listed in Table 5). Using these results, it is possible to narrow the focus of discussion down to two of the four final models. Table 4 suggests that models Control 1 and Base 1 are more likely to follow a normal distribution than models Control 2 or Base 2. Table 5 also suggests that models Control 1 and Base 1 are more reliable than models Control 2 or Base 2. Therefore, the original research questions will be answered in the context of models Control 1 and Base 1.

Question 1: What control variables (if any) most influence student engagement?

Of the two final models used for discussion, only one includes control variables. The reduced model Control 1 uses three control variables for prediction: *Automotive Engineering Technology*, *Chemistry*, and *Program Level*. These are the most influential control predictors for student engagement.

Question 2: What classroom characteristics most influence student engagement?

From the results of reducing the models Base 1 and Control 1, it seems that the *Students' Reluctance to Leave* and *Students' Discussions with Peers* are important in predicting student engagement in the classroom.

Question 3: Does the treatment of missing values affect the models?

Since the models using Data Set 1 (Control 1 and Base 1) seem to be much better than those using Data Set 2 (Control 2 and Base 2), one may conclude that the treatment of missing values does matter. Replacing missing values with the corresponding variable median does not produce models as reliable as simply removing observations with missing values.

Comparing these results with Hoffman's, the two sets of models are not identical, but *Students' Reluctance to Leave* and *Students' Discussions with Peers* are used in all eight final models (the four from Hoffman's paper and the four found here). Also, the variables found influential for both of the Data Set 2 models here (*Amount of Theoretical Focus* and *Instructor Efforts to Connect*) were found in Hoffman's final models. Hoffman's models, however, did not use any transformations on the response variables, whereas here, three of the models required transformation in order to obtain normal residuals. The DPE p-values found for Hoffman's models are generally higher than those found for the models listed here. A higher p-value in this context means a lower chance of rejecting the null hypothesis that the residuals are normal. Therefore, higher p-values are considered better than lower p-values. This becomes less of an issue when we consider that the models with p-values closer to the cut-off of rejecting the null hypothesis (Control 2 and Base 2) were not found to be the best models when comparing all four. The models that performed better under the PRESS and SSE comparison (Control 1 and Base 1) had the highest DPE p-value, reducing concern about the validity of the final results. Hoffman's PRESS and SSE results for the final four models are also quite different from those found here. The differences between the two statistics are extremely similar for all four of Hoffman's models, whereas the differences vary quite drastically for the models found here. There does not seem to be the stark contrast in results found between Data Set 1 and Data Set 2 in Hoffman's that there is here. Overall, the results found with the 'dirty' data and the 'clean' data are similar when it comes to the influential factors found. In both Hoffman's paper and this paper, Base 1 is found to be one of the best two models. Again, the two Base 1 models are not identical, but they do both include *Students' Reluctance to Leave* and *Students' Discussions with Peers*.

In both cases, *Students' Reluctance to Leave* is associated with a positive β-coefficient. This implies that engagement is positively correlated with students staying on task until the end of class. A class where students are often packing up early and "checking-out" would be a sign of low student engagement. Also, in both the Control 1 and Base 1 models, *Students' Discussions with Peers* is associated with a negative β-coefficient. This means one would expect lower student engagement in a class where students are often chatting with one another.

The trouble with making this assumption is that the construct *Students' Discussions with Peers* was created using the two variables E12 and E13 which differ in positivity/negativity. E12 reads: "Do students discuss class material with each other during class?" It could be presumed that class-related discussion among students is a good thing, depending on the context of the class. E13 reads: "Do students chat with each other about non-class topics during class?" This is worded so that a higher scoring is a negative response. The two questions were grouped because it was difficult for observers to distinguish between discussion of class and non-class topics. However, averaging the two variables to create one construct creates confusion in their meaning within the models. Since the goal of this paper was to compare analysis of the clean and dirty data, methods of construct creation similar to those in Hoffman's paper were used here and the two were combined, but different methods of construct creation could be looked at in the future.

The control variable *Program Level* is included in both Hoffman's final models and the models found here, meaning a student's level within in their major or program is important when determining their level of engagement in the classroom. However, there was a difference in coding for the variable *Program Level.* Hoffman used dichotomous dummy variables to create separate variables for sophomore and junior levels, whereas in this paper, one variable was created with a different value for each level. This makes it more difficult to compare results. The results of this paper suggest that the farther along a student is within their program, the lower their engagement because *Program Level* is associated with a negative β-coefficient, which is contrary to what one would expect. This result is similar to results given in another paper by the CTC researchers (Floyd-Smith, et al., 2012) where belonging by program level was investigated. The researchers expected to find students' sense of belonging to increase as they progressed through their program, but this was not the case. Hoffman's results do not allow us to define a sense of directionality because a student's program level was coded to be a nominal variable. We can assume that a student's program level does relate in some way to student engagement from the fact that it is included in both final models.

5. Conclusions and Future Work

Given the results found in this paper, there seems to be a disconnect between the two data sets, Data Set 1 having any missing value observations removed and Data Set 2 having the variable median replace any missing values. Data Set 1 outperformed Data Set 2 at each step of the methodology. However, this was not the case for Hoffman during her experiments. Therefore, one is led to believe that something happened during the cleaning process to change the original data in such a way that treatment of missing values makes a significant difference. The cleaning process may have resulted in more missing values overall, making the need for replacement of missing values more frequent during the creation of the clean Data Set 2. More work must be done to fully understand what happened to create this inconsistency.

Another trouble that arose was the conflicting meaning of grouped variables during construct creation. As mentioned in the discussion section, the construct *Students' Discussions with Peers* was created from two inconsistent variables. One variable associated a higher value with a better response, whereas the other variable associated a lower value with a better response. One solution may be to reverse code one of the variables so the final construct has more meaning, but this has not been done at this point. The cleaning process looked to fix other instances in the data similar to this, but there may still be a need to take another look at the way constructs were created. The method used here is based on expertise from educational theory, but other methods could be used such as factor analysis to create meaningful constructs.

In both Hoffman's paper and this one, attentiveness was used as an indicator of student engagement. However, according to experts, attendance is also a good indicator of student engagement. Therefore, as future work, one could use attendance as the response variable for analysis and compare the results of these two papers. This may require extra work since attendance was not always recorded for each observation in the original data.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DRL-0910143.

The authors thank Wendy Hoffman and Dr. Mezbahur Rahman (Dept. of Mathematics & Statistics, Minnesota State University, Mankato) for their guidance and support of this work. In addition, we thank all members of the "Connections to Community" collaboration, led by university principal investigators: Dr. Denise Wilson, Dr. Melanie Plett, Dr. Tamara Floyd-Smith, and Dr. Nanette Veilleux (National Science Foundation grant numbers DRL-0909817, 0909659, 0909900, and 0909850).

7. References

Allendoerfer, C., Wilson, D., Bates, R., Crawford, J., Jones, D., Floyd-Smith, T., et al. (2012). Strategic Pathways for Success: The Influence of Outside Community on Academic Engagement. *Journal of Engineering Education, 101*(3), 512-538.

Bates, R. A., Martin, J. P., Wilson, D., Plett, M., & Floyd-Smith, T. (2013). Interactive Session: Measuring the Impact of Connection to Community. *120th ASEE Annual Conference & Exposition.* Atlanta: American Society for Engineering Education.

Classroom Observation Instruments. (n.d.). Retrieved November 24, 2013, from University of Minnesota: Center for Teaching and Learning: http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments/

D'Agostino, R. B., & D'Agostino, R. B. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician, 44*(4), 316-321.

D'Agostino, R., & Pearson, E. S. (1973). Tests for Departure from Normality. *Biometrika Trust, 60*(3), 613-622.

Dielman, T. E. (1996). *Applied Regression Analysis for Business and Economics* (Second ed.). Belmont, CA: Wadsworth Publishing Company.

Dielman, T. E. (2005). *Applied Regression Analysis: A Second Course in Business and Economic Statistics* (4th ed.). Belmont, CA: Thomson Learning, Inc.

Floyd-Smith, T., Wilson, D., Campbell, R., Jones, D., Bates, R., Peter, D., et al. (2010). A Multi-Institutional Study of Connection, Community and Engagement in STEM Education: Conceptual Model Development. *American Society for Engineering Education*.

Floyd-Smith, T., Wilson, D., Carlson Jones, D., Plett, M., Bates, R. A., & Veilleux, N. M. (2012). Investigation of Belonging for Engineering and Science Undergraduates by Year in School. *2012 ASEE Annual Conference.* San Antonio: American Society for Engineering Education.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, (pp. 82-88).

Hoffman, W. L. (2013, May). Identifying Influential Variables of Student Academic Engagement (Master's Thesis). Mankato , MN: Minnesota State University, Mankato.

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review, 55*(2), 163-172.

Kim, M. J., Wilson, D., Burpee, E., Allendoerfer, C., Hoffman, W., & Bates, R. (2012). Classroom Influences on Engineering Student Engagement (Manuscript in preparation).

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models* (Fourth ed.). New York, NY: McGraw-Hill/Irwin.

Plett, M., Carlson Jones, D., Crawford, J. K., Floyd Smith, T., Peter, D., Scott, E. P., et al. (2011). STEM Seniors: Strong Connections to Community Are Associated with Identity and Positive Affect in the Classroom. *2011 ASEE Annual Conference & Exposition.* Vancouver.

Rahman, M., & Pearson, L. M. (2008). A Note on the Maximum Likelihood Box-Cox Transformation Parameter. *Journal of Probability and Statistical Science*, 155-168.

Rahman, M., & Wu, H. (2013). A Note on Normality Tests Based on Moments. *Far East Journal of Mathematical Sciences*, 273-282.

Seymour, E. (2002). Tracking the Processes of Change in US Undergraduate Education in Science, Mathematics, Engineering, and Technology. *Science Education*, 79-105.

Wilson, D., Allendoerfer, C., Kim, M. J., Burpee, E., Bates, R. A., Floyd-Smith, T., et al. (2013). STEM Students Outside the Classroom: The Role of the Institution in Defining Extracirricular Activity. *120th ASEE Annual Conference & Exposition.* Atlanta: American Society for Engineering Education.

Zhang, J., & Wu, Y. (2005). Likelihood-ratio tests for normality. *Computational Statistics & Data Analysis, 49*, 709-721.

7. Biographies

Sarah Painter

Sarah Painter graduated from Minnesota State University, Mankato with a B.S. in Mathematics and French in 2013. She worked under Dr. Rebecca Bates as part of her STEM education research. Previous project include statistical analysis of STEM student learning styles and self-efficacy. She plans to return to MSU in pursuit of an M.A. in Mathematics in the fall of 2014.

Dr. Rebecca Bates

Dr. Rebecca Bates is a professor in the Department of Integrated Engineering. Her PhD in Electrical Engineering is from the University of Washington where she worked to develop computer modeling of pronunciation to improve automatic speech recognition. She has a degree in theological studies from Harvard Divinity School, an M.S. in Electrical Engineering from Boston University and a B.S. in Biomedical Engineering from Boston University. Current research projects include the Speech Recognition Virtual Kitchen, a toolkit to support research and education in the field of speech recognition and a multi-institution effort examining the impact of connection to community on student engagement in undergraduate STEM majors.

Appendix A. Observation questions and associated variables

E1. Do students maintain eye contact with the instructor or maintain attention to the instructor while he/she is speaking to the class?

E2. Do students remain awake and alert?

E3. Do students express body language indicative of being *bored or disinterested*/*engaged or interested*? (inconsistent question)

E4. Do students avoid alternative activities (e-mail, newspaper, texting, etc.)?

E5. Do students answer questions when asked by the instructor?

E6. Do students ask questions of their own initiative?

E7. Do students ask follow-up questions/create dialogue with the instructor?

E8. In small group activities, do students stay on task?

E9. In small group activities, do students actively participate?

E10. Are students on-time to class?

E11. Do students *begin to pack up*/*wait to pack up* before the "bell rings"? (inconsistent question)

E12. Do students discuss class material with each other during class?

E13. Do students chat with each other about non-class topics during class?

I1. Are visual aids (including use of white board or chalk board, slides) well organized?

I2. Is the teacher interested, enthusiastic, and engaged in teaching?

I3. Does the instructor use student names?

I4. Does the instructor use humor appropriately?

I5. Does instructor not embarrass or belittle students in any way?

I6. Does the instructor ask questions of students and pause for responses?

I7. Are questions asked by instructor amenable to response from students?

I8. Does the instructor use group activities, think/pair/share, or similar active learning techniques?

I9. Does the instructor have eye contact with students?

I10. Does timing of classroom activities consider attention spans?

I11. Does instructor use different types of explanation for the same problem?

I12. Did the opening gain the class' attention? Did it establish rapport?

I13. Did the opening outline the topic and purpose of the lecture?

I14. Is the delivery paced to students' needs?

I15. Could the instructor be seen and heard?

I16. Were key points emphasized?

I17. Was the lecture stimulating and thought provoking?

I18. Is the instructor available before or after class?

I19. Does the instructor relate class to course goals, students' personal goals, or societal concerns?

I20. Is the instructor paying attention to cues of boredom, confusion?

C1. Are the class web pages or other class documents well organized and easy to navigate?

C2. Do the web pages and/or syllabus provide complete logistics for the course?

C3. Does the syllabus offer "active" verb learning objectives?

C4. Is the classroom environment comfortable for learning (lighting, temperature, chairs, etc.)

C5. Is the class oriented toward theory/basic science?

C6. Is the class oriented toward applications and examples?

C7. Is the class oriented toward design?

DOI: 10.56816/2378-6949.1050