AC 2012-3757: USING HIGH SCHOOL AND DISTRICT ECONOMIC VARIABLES TO PREDICT ENGINEERING PERSISTENCE

Dr. Marisa K. Orr, Purdue University

Marisa K. Orr is a Postdoctoral Researcher in the School of Engineering Education at Purdue University, where she is part of the core MIDFIELD team. MIDFIELD is the Multiple-Institution Database for Investing Engineering Longitudinal Development that includes academic transcript data from eleven public universities from 1988-2009. The focus of Orr's research is on developing new metrics and models to enhance our understanding of students, institutions, and policies. Orr earned her Ph.D., M.S., and B.S. in mechanical engineering from Clemson University, as well as a certificate in engineering and science education.

Nichole M. Ramirez, Purdue University

Nichole M. Ramirez is a graduate student in the School of Engineering Education at Purdue University. She received her B.S. in aerospace engineering from the University of Alabama in 2010 and will complete her M.S. in aviation and aerospace management in May 2012. She is a recipient of the Purdue Doctoral Fellowship and currently serves as Committee Chair of the Engineering Education Graduate Student Association. In addition to socioeconomic research, she is also interested in studying ways to integrate aerospace engineering and aviation technology education.

Dr. Matthew W. Ohland, Purdue University, West Lafayette

Matthew W. Ohland is Associate Professor of engineering education at Purdue University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by more than $11.6 million from the National Science Foundation and the Sloan Foundation and his team received the William Elgin Wickenden Award for the Best Paper in the Journal of Engineering Education in 2008 and multiple conference Best Paper awards. Ohland is Past Chair of ASEE's Educational Research and Methods division and an At-large member the Administrative Committee of the IEEE Education Society. He was the 2002-2006 President of Tau Beta Pi.

Valerie Lundy-Wagner, New York University

Valerie Lundy-Wagner is is currently an Assistant Professor and Faculty Fellow in higher and post-secondary education at the Steinhardt School for Culture, Education, and Human Development at New York University. Her research focuses on how institutions can better support degree completion, with a special interest in demography (i.e., ethnicity/race, gender, and social class), engineering/STEM, and minority-serving institutions. She earned her doctorate in higher education at the University of Pennsylvania, master's at education at Stanford University, and B.S. in civil and environmental engineering at UCLA.

Using High School and District Economic Variables to Predict Engineering Persistence

Abstract

Prior research has shown that Peer Economic Status (PES), a socioeconomic indicator based on a school's free lunch participation, is predictive of enrollment in engineering, first-year GPA, and engineering degree completion. In that study, PES was calculated as an average over the entire time period (1987-2004). To further explore the utility of this variable two new time-variant forms will be used, computed at the school-level and the district-level. Academic variables are drawn from the Multiple Institution Database for Investigation of Engineering Longitudinal Development (MIDFIELD) database and high school codes are used to link data from the National Center for Education Statistics (NCES).

The time-variant PES is calculated from the four years each student is expected to have been in high school. Additionally, a new algorithm for the treatment of missing values is utilized. The district economic status (DES) is computed in a similar fashion. A series of logistic regression models is used to determine the impact of school- and district-level economic status variables on six-year degree completion. Results show that the time-averaged measures are stronger indicators of engineering persistence than the time-variant measures and that school-level variables are better predictors than district-level variables. Additionally the importance of context in interpreting socioeconomic variables is highlighted.

Introduction

Socioeconomic status (SES) continues to be a major issue in educational equity, diversity, and policy research. Access to higher education and academic achievement is an especially important issue in engineering education. As the United States continues to support STEM initiatives in a highly competitive global market, it is critical that we understand the barriers students face in obtaining an engineering degree. With the goal of informing future research, we seek to answer the questions:

1. Does using a time-variant measure of economic status better predict engineering persistence than a time averaged measure?
2. How do school and district-level measures of economic status compare in predicting engineering persistence?

We hypothesize that the specificity of a time-variant measure will make it a better predictor of engineering persistence and likewise, a school-level measure will be a better predictor than a district-level measure. It is important to explore the relationship between school and district free lunch measures because although we expect the school measures to be the best, many other socioeconomic variables are only available at the district level.

Socioeconomic Status and Academic Achievement

Low-income students are not only less likely to attend college; they are also less likely to attend selective institutions than their high-income counterparts [1, 2]. The limited resources available to these students create additional road blocks to upward social mobility. In K-12 schools with high concentrations of poverty, teachers are less likely to have master's degrees [3] and equipment and internet access are lacking [4]. Perry and McConney show that "increases in the mean SES of a school are associated with consistent increases in students' academic achievement and that this relationship is similar for all students regardless of their individual SES" [5]. Furthermore, Unnever found that a school district's SES may have an even greater impact on student achievement than the educational resources of that district [6].

Low-income students often have higher unmet financial need and are less academically prepared than their higher-income peers, resulting in a higher probability of not completing a college degree [7, 8]. Although there have been many persistence studies in engineering education, research on the impacts of SES is limited. The work of Fenske et al. is one work that does integrate financial aid, gender, race/ethnicity, and academic major and implies that STEM majors often come from higher SES strata and have higher measures of academic achievement [9]. It is evident that research suggests a correlation between SES and access to postsecondary education as well as persistence in engineering.

Free lunch literature

Research shows that free and reduced lunch eligibility is a valid predictor of student achievement in the United States [10] as well as abroad [11]. Free Lunch is part of the National School Lunch Program administered by the U.S. Department of Agriculture. Children from eligible households receive a school lunch at no charge. To be eligible, a household income must be less than 130% of the poverty guidelines published by the Secretary of agriculture each year [12]. The 2011 poverty guideline for a household of four is $22,350 [13], making children from households with income less than $29,055 eligible for free lunch.

Using free and reduced lunch enrollment as an indicator of both school and family poverty status, Caldas and Bankston show that there is a negative correlation between individual academic achievement and both individual and school poverty statuses. School poverty status was only slightly less correlated than individual poverty status [14]. Previous work linking academic variables from the Multiple Institution Database for Investigation of Engineering Longitudinal Development (MIDFIELD) database to free lunch data from the National Center for Education Statistics (NCES) indicated that free lunch eligibility of a school, peer economic status (PES), was a significant predictor of individual variables. Furthermore, the study concluded that students from low-income schools are at a disadvantage when it comes to engineering enrollment, academic achievement, and six-year graduation [15].

From 1990 to 2010 participation in the National School Lunch Program expanded from 24 to 31.7 million children [12]. An approximately 32% increase over this time period should be taken into account with a time variant economic variable; however, most studies use only a single point in time to compute a measure of school SES [5, 8, 14]. What makes this study unique is the compilation of nearly 20 years of Common Core Data integrated with longitudinal university data, taking into account students' specific high school time period. Researchers acknowledge the need to better understand the barriers economically disadvantaged students face in higher education [7, 8] to inform academic advising and policy making.

Method

The goal of this study is to compare the predictive ability of free lunch variables on two dimensions: time-averaged vs. time-variant, and school-level vs. district-level. The outcome variable is six-year graduation in engineering. This outcome was chosen over engineering enrollment, first-year persistence, and first-year GPA because prior work has shown it to be the most readily explained by a time-averaged school-level free lunch [15]. The study population is drawn from the MIDFIELD database and high school codes are used to link data from the NCES.

MIDFIELD

MIDFIELD includes twenty years of student record data from eleven partner institutions, including four of the ten largest U.S. engineering programs in terms of undergraduate enrollment. This study used only first-time-in-college students because high school codes are often not reported for transfer students. International students were also excluded. Two institutions were excluded because high school codes were not reported for over 40% of their students. For the same reason, academic years 1992/93 and 1993/94 were excluded from another institution. Data were also buffered to include only students with at least six years of institution data available which brought the total number of students to 265,549. Of these, the 50,866 who matriculated in an engineering program are included in our sample

NCES

The National Center for Education Statistics (NCES) maintains a database of all public elementary and secondary schools in the United States. This Common Core of Data (CCD) contains information on participation in the National School Lunch Program collected annually at both the school and district level. Additional fiscal and socioeconomic data are available for each district.

The Common Core of Data [16] is adjusted by NCES to protect privacy. If a school reports all students on free lunch, NCES adjusts the reported value to 95% (1999–2000 and earlier) or enrollment minus 3 (starting in 2000-2001) so that no individual student can be identified as eligible for free lunch [17].

Many schools reported zero enrolled in Free Lunch in the earlier years, probably because the school was not actively participating, or perhaps not reporting it. Due to this pattern, all zeros are counted as missing until a school reports at least one student eligible for Free Lunch. By this definition, more than half the schools are missing free lunch enrollment data each year before 1991.

Variables

All of the economic status variables are computed as 100% minus the percentage of students eligible for free lunch in a particular school or district. Therefore, high values represent high economic strata. It is important to note that the variables are not an indication of a student's household economic status, but rather an indicator of their school or district environment.

The school-level variables, PESavg and PESvar, include all student peers in all grade levels at the school. The district-level variables, DESavg and DESvar, include all schools (elementary and secondary) in the district and hence tend to be lower because students in lower grades are more likely to enroll in the National School Lunch Program. The time-averaged variables (ending in -avg) use a weighted average of all the free lunch and enrollment data available over the 1988-2009 school years. The time-variant variables (ending in -var) are specific to the years a student is expected to have been in high school (the four years before their college enrollment). If no data were available for the four years a student is expected to have been in high school, the value is imputed as the PESavg or DESavg for the high school. Both time-averaged and time-variant variables are computed for students starting college in 1992 or later. This allows four years of free lunch data for the time-variant variables.

The outcome variable, six-year graduation in engineering was chosen based on prior work [15] that showed it was the outcome best predicted by PESavg (referred to in that paper as pSES). This variable indicates whether a student graduates within six-years of their start date in an engineering field.

Descriptive Statistics

Over 90% of the 50,866 engineering matriculants had CEEB codes that could be identified as either public or private. Engineering matriculants from public and private schools had similar rates of six-year graduation in engineering (Table 1), although the group of students with missing or unidentifiable high school codes had a higher graduation rate.   Free lunch eligibility data are only available from public schools, so both private and unknown are automatically considered missing.

Table 1. Engineering Graduation Rates by High School Type

High School Type	Engineering matriculants	Percent of sample	Percent who graduate in engineering in 6 years
Public	42225	83.1%	50.4%
Private	4085	8.0%	50.6%
Unknown	4526	8.9%	53.2%

The school-level and district-level economic variables are missing for 18.5 and 18.1% of the sample, respectively. Economic status data could be missing due to four causes: no valid CEEB high school code was reported by the postsecondary institution; the student attended a private school; the CEEB school code could not be matched to a school in the NCES table; or free lunch counts in the NCES tables are missing or zero for all years in the range 1988-2009. For a fair comparison, only the records with both school and district data are used in the logistic regression analyses. This common subset includes 41,413 MIDFIELD students, with economic data as shown in Table 2.

Table 2. Descriptive Statistics

Variable	Mean	Standard Deviation	25th Percentile (Lower Quartile)	50th Percentile (Median)	75th Percentile (Upper Quartile)
PESavg	85.6	11.3	79.7	88.4	94.2
PESvar	88.8	10.5	84.6	92.0	96.2
DESavg	77.0	14.1	67.7	79.3	88.0
DESvar	79.8	13.8	71.6	83.6	90.3

Analysis

The analysis consists of four logistic regression models that estimate the probability of six-year graduation in engineering. The null model includes only institution. The four experimental models each contain institution, an economic status variable, and the interaction of the two. This allows us to determine how much more variance is explained by including the economic status variable.

Raudenbush and Bryk assert the importance of using hierarchical linear modeling (HLM), or multilevel modeling (MLM), in education research, especially when using variables that are aggregated at a higher level than the outcome variable(s) [18]. In our case, six-year graduation is a student level outcome while PES and DES are variables that are aggregated at the school and district levels, respectively. MLM takes into account the interrelatedness of variables at multiple levels, which violates the assumption of independence in ordinary least squares (OLS) regression [19]. While MLM preserves the structure of multi-level data, it is more complicated in implementation and interpretation. Others provide evidence that MLM may be unnecessary in this study. In Astin and Densons's comparative study between MLM and OLS, the authors conclude that, in most cases, the two methods provide similar results [19]; therefore, we utilize stepwise logistic regression models in this study. They do, however, recommend reducing the alpha level for institutional effects by half to guard against type I errors. Due to this recommendation and the large sample size, a stringent alpha level of 1% was used to reduce the possibility of detecting false differences.

Results & Discussion

We hypothesized that making the free lunch variables time-variant would increase the predictive power by creating more variance and accounting for economic changes over time, however, the results suggest otherwise. The time-averaged school-level variable (PESavg) performed better than any other. As shown in Table 3, all the variables had significant ($p<<0.01$) main effects and interaction with institution. Max-rescaled R-square is an indication of the variance explained by a particular model. The base model, which includes only the institution, explains 1.9% of the variance. Adding PESavg and the PESavg*institution interaction to the model increases the variance explained by 35%. DESavg also outperformed DESvar, suggesting that long term conditions carry more weight than the current conditions. As expected, school-level variables better estimated graduation rates than district-level variables. The district level-variables were however still beneficial, which is important because additional data are available at the district-level.

Table 3. Comparison of Economic Status Variables

Model	Max-Rescaled R-Square	β Estimate (main effect)	Significance (main effect)	Significance (interaction)	Improvement
Institution only	0.0199	-	-	-	-
PESavg	0.0268	0.00786	<0.0001	<0.0001	35%
PESvar	0.0250	0.00677	<0.0001	<0.0001	26%
DESavg	0.0257	0.00697	<0.0001	0.0041	29%
DESvar	0.0245	0.00507	0.0008	0.0003	23%

Based on the results presented in Table 3, PESavg will be used for the remainder of this paper and for future work. The institutional interaction effect is significant, which suggests that the effect of a student's high school or district environment is moderated by their institution. Six-year graduation rates in engineering vary by institution from 23% to 56% and the mean PESavg value for an institution ranges from 73-90. Figure 1 illustrates the predicted probability of graduation in engineering within six years for students at each institution by PESavg.

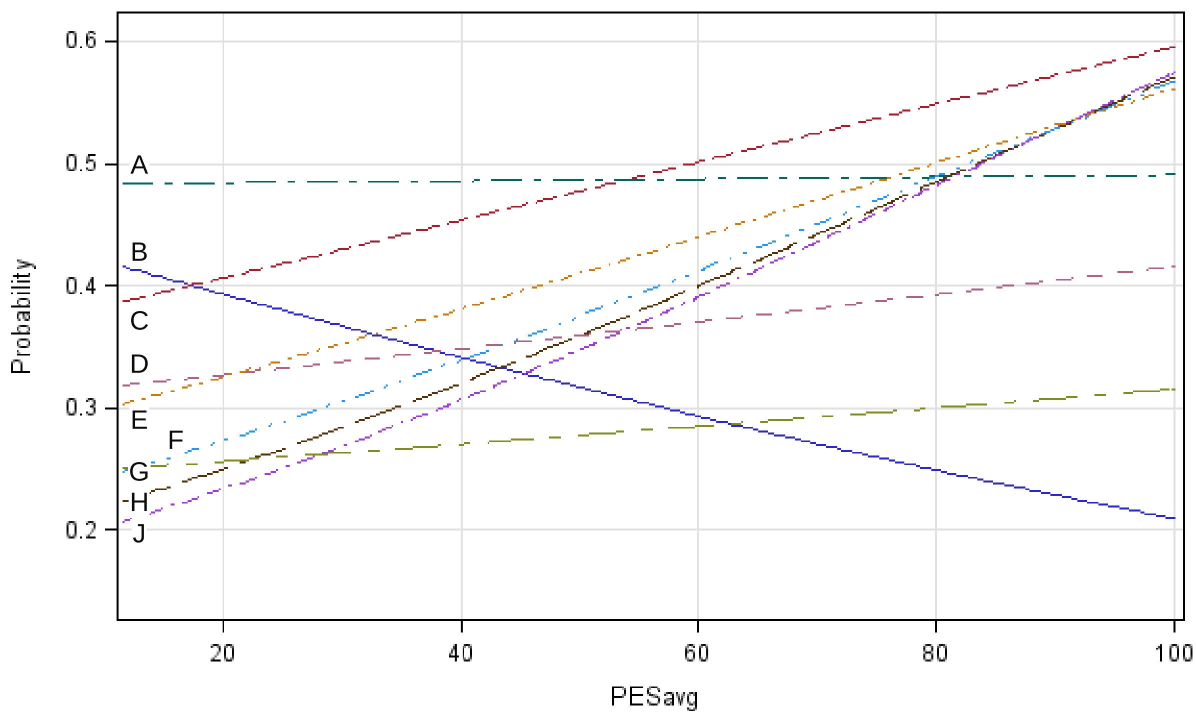

Figure 1. Predicted probability of six-year graduation in engineering by institution

At institutions C,E,F,H and J, students with high values of PESavg are most likely to graduate in engineering, while other institutions show a more neutral effect (A, D, and G). Only one institution, B, showed that PESavg has a negative effect on six-year graduation. Examination of the confidence limits in Table 4 reveals that this particular effect is not significant within 99% confidence limits. Still, this is distinct from expectation, and is being explored further with administrators at that campus. Because this effect is unique to a single institution, we cannot share our speculations without revealing the institutional identity, yet this particularly highlights the importance of context in interpreting socioeconomic effects.

Table 4. Odds Ratio by institution for PESavg in increments of 10

Institution	Odds Ratio for increase of 10 in PESavg	Lower Confidence Limit (99%)	Upper Confidence Limit (99%)
J	1.204	1.145	1.267
H	1.189	1.111	1.272
F	1.169	1.087	1.257
E	1.130	1.057	1.207
C	1.100	1.045	1.159
D	1.049	0.952	1.155
G	1.037	0.860	1.250
A	1.003	0.911	1.105
B	0.894	0.739	1.082

The odds ratios presented in Table 4 allow us to compare the odds of high- and low-PES students at the same institution. Increments of 10 are chosen to illustrate this point in a meaningful way (a 1% difference in free lunch eligibility is difficult to conceptualize). As the name suggests, the odds ratio (OR) is a ratio of odds. The denominator is the odds of a student of PESavg=x graduating in engineering and the numerator is the odds of a student of PESavg = x+10 graduating in engineering. So if students of all socioeconomic backgrounds have an equal chance of success, the odds ratio will be 1. An OR of more than 1 indicates that students of high SES have more favorable odds of graduation than those of low SES. The shaded institutions have OR confidence intervals that include 1, indicating that they are not significant. The range of significant OR values indicates that, at those institutions, a 10 point shift in PESavg can influence graduation likelihood by 10% to 20%.

Limitations

Although there are limitations to the conceptualization and construction of the PES variable, the research presented suggests that economic status is an important factor to consider in postsecondary outcomes. Economic status is a significant predictor in models of degree completion at colleges that produce large number of engineering bachelor's degree recipients. Although this research is limited by the types of institutions included in the model (large public universities with large engineering programs), it highlights the contribution that a student-level pre-college measure of SES affords models of postsecondary outcomes.

Even when PES is included, the model still captures less than 3% of the variance present in the data. Thus, it is critical to recognize that the value of the model is in understanding the effects of the model's variables on the behavior of groups of students. It would be unwise to use this model to predict the likelihood that a particular student will graduate in engineering. One promising aspect of this research is that PES reflects the cultural and economic resources of public schools; unlike race or gender, school environments can be changed. Future research along these lines can help identify high school characteristics that help or hinder students' postsecondary opportunities.

Conclusions and Future Work

Contrary to our hypothesis, a variable based a long-term average of free lunch participation is a better predictor of six-year graduation in engineering than one focused on the period a student was expected to have been in high school. In alignment with our expectations, variables aggregated at the district level were significant, but could not explain as much variance as the school-level variables. School-level variables are more representative of a student's experience. District-level variables, have the statistical advantage of a greater range of values, but lack specificity with respect to the student's experience. The fact that district-level variables are significant is still important, because many more variables are available at the district level, which could improve this predictive model as those additional variables are incorporated into the definition of DESavg. Our next steps will be to begin conditioning other socioeconomic variables such as expenditures per student and census data such as household income, education, and employment levels. Eventually, these variables will be used in a cross-classified multi-level model to further understand the effects of school/district socioeconomic variables on college performance and the moderating or exacerbating effects of postsecondary institutions. Such a model should be more powerful and more authentic because it will partition the variance in a way that matches the natural structure of the data. Whereas the max rescaled R-squared for any of these models is low, this is largely because a significant amount of the variability resides with individual students and cannot be explained. The variables still have clear predictive validity.

The validity of the PES variable suggests that students do face economic barriers in higher education; more specifically, they encounter obstacles to graduation in engineering. It is important to remember that mitigating the impacts of those barriers may be achieved by establishing equity in education. Our results indicate that educational equity is possible and is already occurring at some schools.

This study is a preliminary step in learning about practical ways to research the effect of socioeconomic status. A better understanding of this phenomenon can provide guidance to policymakers and educators for establishing a more equitable education system, both in how resources are allocated at the k-12 level and in how students are encouraged and supported in higher education. While some institutions favor students from high-SES backgrounds, others have fostered a climate in which all students have a fair chance at success. By identifying the characteristics that make certain institutions more equitable, we can enhance opportunities for students of all backgrounds.

References

1. Carnevale, A.P. and S.J. Rose, *Socioeconomic Status, Race/Ethnicity and Selective College Admissions*, in *America's Untapped Resource: Low-Income Students in Higher Education*, Richard D. Kahlenberg, Editor 2004, Century Foundation Press: New York. p. 106, Table 3.1.
2. *Socioeconomic Diversity: Advantage for Wealthy Students*. The Chronicle of Higher Education, 2010.
3. Aud, S., et al., *The Condition of Education 2010*, National Center for Education Statistics. 2010.

4. May, G.S. and D.E. Chubin, *A Retrospective on Undergraduate Engineering Success for Underrepresented Minority Students.* Journal of Engineering Education, 2003. **92**(1): p. 27-39.
5. Perry, L.B. and A. McConney, *Does the SES of the School Matter? An Examination of Socioeconomic Status and Student Achievement Using PISA 2003.* Teachers College Record, 2010. **112**(4): p. 1137-1162.
6. Unnever, J.D., A.C. Kerckhoff, and T.J. Robinson, *District Variations in Educational Resources and Student Outcomes.* Economics of Education Review, 2000. **19**: p. 245-259.
7. King, J.E., *Crucial Choices: How Students' Financial Decisions Affect Their Academic Success*2002, Washington, DC: American Council on Education.
8. Walpole, M.B., *Socioeconomic Status and College: How SES Affects College Experiences and Outcomes.* The Review of Higher Education, 2003. **27**(1): p. 45-73.
9. Fenske, R.H., J.D. Porter, and C.P. DuBrock, *Tracking Financial Aid and Persistence of Women, Minority, and Needy Students in Science, Engineering, and Mathematics.* Research in Higher Education, 2000. **41**(1): p. 67-94.
10. Knoeppel, R.C. and C.A. Brewer, *Education Reform, Equal Opportunity and Educational Achievement: Do Trend Data Adequately Report Progress?* Education Policy Analysis Archives, 2011. **19**(10): p. 1-28.
11. Lubienski, S.T. and C.C. Crane, *Beyond Free Lunch: Which Family Background Measures Matter?* Education Policy Analysis Archives, 2010. **18**(11): p. 1-39.
12. Child Nutrition Programs. *National School Lunch Program*. 2010 [cited 2011 September 22]; Available from: http://www.fns.usda.gov/cnd/lunch/AboutLunch/NSLPFactSheet.pdf.
13. Sebelius, K., *Federal Register*, Department of Health and Human Services. 2011. p. 2.
14. Caldas, S.J. and C. Bankston, *Effect of School Population Socioeconomic Status on Individual Academic Achievement.* The Journal of Educational Research, 1997. **90**(5): p. 269-277.
15. Orr, M.K., N.M. Ramirez, and M.W. Ohland *Socioeconomic Trends in Engineering: Enrollment, Persistence, and Academic Achievement*. Proc. American Society for Engineering Education Annual Conference, 2011.
16. National Center for Education Statistics, *Public Elementary/Secondary School Universe Survey Data*.
17. Sable, J., J.M. Thomas, and J. Sietsema, *Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2004–05*, National Center for Education Statistics, Institute of Education Sciences, and U.S. Department of Education. 2007: Washington, DC.
18. Raudenbush, S. and A.S. Bryk, *A Hierarchical Model for Studying School Effects.* Sociology of Education, 1986. **59**(1): p. 1-17.
19. Astin, A.W. and N. Denson, *Multi-Campus Studies of College Impact: Which Statistical Method is Appropriate?* Research in Higher Education, 2009. **50**: p. 354-367.