AC 2012-4517: SECTIONALITY OR WHY SECTION DETERMINES GRADES: AN EXPLORATION OF ENGINEERING CORE COURSE SECTION GRADES USING A HIERARCHICAL LINEAR MODEL AND THE MULTIPLE-INSTITUTIO DATABASE FOR INVESTIGATING ENGINEERING LONGITUDINAL DE-VELOPMENT

Mr. George D. Ricco, Purdue University, West Lafayette

George D. Ricco is a doctoral student in Purdue University's School of Engineering Education. He previously received an M.S. in earth and planetary sciences studying geospatial imaging and a M.S. in physics studying high-pressure, high-temperature FT-IR spectroscopy in heavy water, both from the University of California, Santa Cruz. He has a B.S.E. in engineering physics with a concentration in electrical engineering from Case Western Reserve University. His academic interests include longitudinal analysis, visualization, semantics, team formation, gender issues, existential phenomenology, and Lagomorph physiology.

Mr. Noah Salzman, Purdue University, West Lafayette

Noah Salzman is a graduate student in engineering education and mechanical engineering at Purdue University. He received his B.S. in engineering from Swarthmore College, and his M.Ed. in secondary science education from University of Massachusetts, Amherst. He has work experience as both an engineer and taught science, technology, engineering, and mathematics at the high school level. His research focuses on the intersection of pre-college and undergraduate engineering programs.

Mr. Russell Andrew Long, Purdue University
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.

Sectionality

or

Why Section Determines Grades: an Exploration of Engineering Core Course Section Grades using a Hierarchical Linear Model and the Multiple-Institution Database for Investigating Engineering Longitudinal Development

Introduction

Grades, how they are earned, and the institutional impetuses that drive them, are an issue of central importance in the engineering discipline.[1-4] How grades are earned, how different institutions address grades and grade inequities, how instructional practices and policies affect grades, and other grading notions have been studied widely in engineering education.[5-8] The effect of faculty on student grades, while studied,[9] has not been probed as extensively within engineering education using a hierarchical linear model (HLM).

One of the great, open questions in engineering education is *whether or not the section makes a difference in a student's grade.* In other words, the effect of *sectionality* on grades to a large extent is unknown. Sectionality combines instructor effects, effects related to time-of-day of instruction, effects related to any tendency for students to coordinate their enrollment, and other effects. Experience and anecdotal evidence suggest that sectionality affects grades, but large-scale empirical studies of this phenomenon do not exist. Due to the inherent structured nature between course sections and students, standard linear regression models do not offer a robust solution to probing longitudinal systems containing multilevel variables. Hierarchical Linear Models (HLMs) provide a robust solution to studying nested or hierarchical systems when compared with standard regression techniques. We constructed a simple HLM to probe inter-section and intra-section variability in grades within the Multiple Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) by the calculation of intraclass correlation coefficient (ICCs).[10, 11] We then examined grades from three sets of courses endemic to the first year engineering experience: the first chemistry course; the first calculus course; and the first physics course. Our preliminary results indicate that the choice of a HLM to analyze our longitudinal database is correct due to strong variability in grades explained by the high intraclass correlation coefficient (ICC) for most of our MIDFIELD institutions across all three course types analyzed.

Basics of HLM

The world of hierarchical linear models (HLMs) is an expansive space, and fundamental texts in the discipline describe it in a detail beyond this paper.[12-14] HLM is known by many names in different fields: the most popular being hierarchal linear models;[15] multilevel models, (first coined by those in the social sciences);[16] generalized linear mixed models, (a name favoured by economists);[7] nested models;[17] mixed models or mixed effects models, (used by biometrics researchers); random coefficient models;[18] random effects models;[19] random parameter models;[20-23] split plot models; covariance components models, (as used by those in statistics); and more.

While sparsely used within the field of engineering education, HLMs harken back to the work of Charles Hendersen, a professor of animal sciences at Cornell University, and a few of his contemporaries.[24] As a modeling specialist, he developed different types of methodologies to

predict breeding outcomes related to butterfat production in cows. In his search for more precise calculations of variances and covariances, he inadvertently developed the HLM that is ubiquitous across academic disciplines today. While contemporaries of his did indeed deal with nested data at the time, Henderson developed the first statistically valid methodology for addressing different variance parameters.

HLMs are models that address data sets organized into hierarchical structures.[11, 15] The three, fundamental classes of HLM implemented across disciplines are: first, the growth curve model, which usually has a time variant structure where observations are nested within individuals; second, the within-person variation model, which proves useful for monitoring repeated measurements for one person or entity within a time invariant structure; and third, the clustered observations model, wherein there are tiered hierarchies such as students within sections within courses or, say, workers within factories within corporations.

The fundamental assumption of HLM is that variance of parameters occurs at more than one level. To organize and analyze data sets in such a hierarchical manner has many distinct advantages: first, acknowledges and addresses the interrelatedness of observations; second, it treats longitudinal data sets in a more natural way than cross-sectional analysis;[25] third, unlike ANOVA variants, HLM does not require the same data structure for all members in any set, so not every member in any set has to have an outcome or measurement at the same interval; fourth, it acknowledges potentially holistic effects such as the impetuses of policy changes at higher levels upon lower levels, (such as institutional policies upon departments, and department policy changes upon instructors, and instructor policy changes upon students);[15] and fifth, for every level in a HLM, a variant of the same predictor can be used, (ex., in a class-student database, student SAT can be used at the student level and average student SAT can be used at the class level).

Caveats of HLM

The plethora of classic HLM papers in the literature can befuddle the novice researcher; however, there are a handful that demonstrate the fundamentals of HLM.[10, 11, 15, 26, 27] Coleman et al's classic paper details a HLM studying cognitive outcomes between private and public schools in the High School and Beyond study.[28] They showed a difference between cognitive outcomes in time between public and private schools, but that result could be easily attained using linear regression. What makes Coleman et al's work historic, is that their HLM showed that private schools engendered a greater, positive change in cognitive values for students who started off at a lower value upon matriculation. Their work illustrates the power of HLM; when used to explore differences in slope between hierarchical sets, HLM can uncover relationships standard linear regression techniques cannot.

Of great concern to the implementation of HLM to longitudinal data in engineering education is the validity of model choice.[25, 29, 30] Although differences between large model categories can readily be explained, (such as those only using first-level variables and those containing second-level variables), more subtle differences, such as the introduction of a new or deletion of a time construct can beget differences. In such time variant models, colloquially referred to as *growth curve models* by HLM researchers, Morrell et al.'s research provides an example of avoiding such a quagmire. [29] By investigating in both a visual and statistical manner, Morrell et al. demonstrate the importance of considering how HLM time measurements are implemented. Specifically, they compare a growth curve model based on the first age of patients, and then introduce a "follow-up" patient time variable, leading to significantly different results. Their conclusion notes that implementing another time variable allowed them to compare and contrast a true, longitudinal model with a more cross-sectional one.

Whereas Morrell et al.'s work warns us of the folly inherit to considering a specific model choice, Astin and Denson's work more generally summarizes the need for more capable methodologies to describe the intricacies of longitudinal data sets.[30] Careful to stress the differences between ordinary least squares (OLS) regression and HLM, Astin and Denson explore a range of variables predicting the future political leanings of over 8,000 college students using OLS and HLM models. While Astin and Denson conjecture that HLM models yield more significant results with variables endemic to institutional-level effects, whereas those pertaining more to path analysis are better served by OLS models, they do not observe any such differences in their study. The key to Astin and Denson's work is that they show in clear terms that one can use two seemingly disparate models and receive strikingly similar results even when using multiple first-level (gender, SES, SAT, etc) and second-level effects (religious affiliation of the college, faculty political orientation, etc).

Why Sectionality: Instructor influence on student achievement

The effect of teachers on student performance is well known and researched in K-12 education.[31] Sanders & Rivers found that a student's sequence of teachers could affect achievement by as much as 50 percentile points. Also, teacher training and background have also been shown to influence student achievement.[9] A full review of the effect of teachers on student achievement is beyond the scope of this paper, but is clearly an established phenomenon at the K-12 level.

Several studies in higher education have demonstrated the effect of faculty on student achievement. Within the engineering context, Vogt (2008) found that faculty distance (defined as courses taught in a large lecture format with limited opportunities to interact directly with the professor) had a strong negative influence on self-efficacy, academic confidence, and GPA.[6] These effects can be particularly pronounced among female students, who tend to report greater numbers of negative interactions with professors and a corresponding loss of academic confidence.[32] Further evidence of the effect of a faculty member on student success is described

by the mismatch of the faculty member's teaching style with the prevailing learning style of the students in the class. Such a mismatch can have serious consequences.[8] The benefits of cooperative learning, peer learning, and active learning are well established in the literature, so the extent to which individual faculty use those instructional methods will have an effect on the success of the classes that they teach. The collective performance of a class of students will also be affected by peer group effects that are beyond the influence of the faculty member's instructional choices. More generally, research on the impact of student-faculty interaction describe another mechanism that will have the effect of creating section-level variability in class performance.[33]

In addition to affecting individual student achievement, significant variance in grade inflation can also be attributed to individual faculty members. For example, in a study of two decades worth of data, Jewell et al. (2011) found that instructor effects were responsible for approximately 40 percent of the University of North Texas's grade inflation over the sample period.[7]

Recent ASEE papers using HLMs or HLM-like models

Borchers and Hee present the notion that the hierarchal nature of the data structures within their entrepreneurship program lend themselves to analysis via HLM.[34] Although none of their work presented within this paper contains analyzed data, they succinctly describe a methodology for addressing hierarchical data using a nested relationship as we have in this current research. They accurately describe the nature of structures such as "peer group, classroom, grade level, school, school district, state, and country,"(ibid, p.2) that are not easily treated as first level; thus, furthering the need for a HLM or a "holistic" (ibid, p.2) analysis.

Borchers and Hee agree with our treatment of a HLM system and the MIDFIELD database in general, in that a preponderance of qualitative impetuses effect something as simple as student placement in a particular section. In other words, no student *choice* is completely random, as many statistical methods assume. In MIDFIELD, our observations have led us to conclude in multiple instances that *institutional policies*, policies that differentiate between institutional cultures, can dominate the statistical evaluation of any student outcome variables.[35] Their proposed methodology overlaps with our current analysis is in the treatment of classroom variance. As they correctly assert, simply "bringing down" (Borchers and Hee, 2011, p.7) classroom variables without a HLM structure can have the deleterious effect of attenuating calculated variability by 80-90%.

Beyond simply expressing institutional level effects on student performance, Padilla et al. note in their 2005 paper the importance of eliminating aggregation bias and misestimated standard errors that occur when researchers ignore the nested structures inherent in HLM. [36] The treatment of HLM in light of aggregation errors by Padilla is elegantly explained as "when an explanatory variable can take on different meanings." (ibid, p.) HLM measures mean values at

each level, so while student grade averages within a section have one meaning, their average at the institutional level has another meaning. Padilla et al. noted a 19% variance in grades based solely on institutional differences.

Within engineering design research, the only recent use of HLM is the work of Lawanto. Using HLM to evaluate self-appraising or self-managing or level of difficulty would prove indispensable to constructing new protocol. Their paper uses a HLM model to evaluate the relationship between two research questions, although their results are not explicit enough to provide serve as a primer for future research. [37]

Faculty training, such as time in industry, [38] has also been recently discussed within our field. Harper and Terenzini's work used a level-2 model to explore the relationships between faculty experience and student participation in co-curricular activities. Unlike many analyses that use models to prove a statistical correlation, their work succeeds by demonstrating that some common factors such as gender, SAT, and high school GPA are statistically insignificant in engendering co-curricular comparison compared to the engineering programs and faculty themselves.

While not to be confused with HLM, hierarchical regression modeling can be successfully used in place of a HLM as another recent work discusses.[39] When the intra-class correlation function approaches a low number, (traditionally around 3% or less depending on various effects,) a multi-level model can be abandoned as long as the researcher is comfortable that the convergence of the resulting variability.

Methodology

A note on our data set and MIDFIELD in general

The MIDFIELD database contains records for 701,190 first-time-in-college students matriculating in any major at participating institutions.[35, 40] The initial population for this study consisted of 137,071 first-time-in-college (FTIC) students who ever matriculated in engineering at one of nine of our MIDFIELD institutions in 1988 and later, for which data is available. We tallied all of the instances when students enrolled in a first semester/quarter, core chemistry, calculus, or physics class required for engineering majors that contained section data. We further pared down this population by removing students who repeated a course and received zero credits as per institution policy and cleaning up any remaining, erroneous section and grade data. The net contention of these routines yielded a data set of 161,456 instances of student who ever declared an engineering major and their grades within three, core course sections.

The MIDFIELD schools are all public institutions and are mostly located in the southeastern United States, yet their size and diversity help make the results generalizable. These partner institutions have larger overall enrollment and engineering programs than average compared to the more than 300 colleges with engineering programs. The partners include six of

the fifty largest U.S. engineering programs in terms of undergraduate enrollment, resulting in a population that includes more than 1/10 of all engineering graduates of U.S. engineering programs. MIDFIELD's female population comprises 22.1 percent of students, which aligns with national averages of 20 percent from 1999-2003[41] and 22 percent in 2005.[42]

The construction and Explanation of a simple HLM

The simplest way to explain HLM to the researcher adept in regression techniques is to write a linear regression formula. For instance, the equation for a line in linear regression is,

$$Y_i = \beta_0 + \beta_1 X_i, + e_i$$

Obviously, this is the equation for a line and hence the name is *linear* regression. Here, Y_i is the dependent variable (or criterion), β_0 is the intercept, β_1 is the slope, X_i is the predictor variable, and e_i is the residual (or sometimes as it is colloquially called, the *error*).

In HLM, a standard model equation looks like,

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_i, + e_{ij}$$

The only difference between our linear regression model and our HLM here is that the HLM model has the added j subscript, which is our nested or clustered unit factor. This means that for every value of our hierarchy there is an outcome (or dependent variable), intercept, slope, and residual. Another way of putting this is that every classroom has its own outcome, intercept, slope, and residual values determined by the students in that classroom.

For our results presented within this paper, we utilized the simplest form of HLM, which is called the *intercept-only model*,

$$Y_{ij} = \beta_{0j} + e_{ij}$$

Because the *intercept-only model* utilizes only intercepts, it is a simple and ideal tool to probe variability within hierarchical structures.[13, 14] The intercept-only model is also called the *null model* or the *empty model* or the *fully unconditional model*.[15]

The most important calculation to undertake when employing the intercept-only model is the intra-class correlation (ICC). The ICC is a ratio that tells the researcher the degree to which the variability discussed in his or her model lies within one structure or another. For instance, where the ICC is given by,

$$ICC = \rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{UN(1,1)}{UN(1,1) + e_{ij}}$$

The factor τ_{00} (or UN(1,1) within SAS models) is the variability between levels and σ^2 (or residual within SAS models) is the variability within levels. In our case for this preliminary model, we have the variability of grades between course sections, and the variability of grades

within those sections. Here, a value of 0.15 would mean that 15% of the total variability of our intercept-only model lies between course sections. Where the ICC falls under 3%, it has traditionally been thought that this threshold indicates that HLM may not gleam useful results and the data are better suited to analysis using traditional linear regression.

Results

After performing an intercept-only analysis on the first core chemistry, calculus, and physics courses with section data, we produced the following results in Tables 1 (calculus I), 2 (chemistry I), and 3 (physics I). Our institutions have been randomly assigned a number that is the same for all three tables. Remember, that UN(1,1) is standard notation for τ_{00}, and in the intercept-only that is the between-level variability. Here, that means UN(1,1) is the variability between course sections. The residual value here is e_{ij} or σ^2, which is the variability within course sections. The intraclass correlation coefficient is calculated as explained earlier in this paper and reported in all three tables. Finally, the convergence factor, usually designated by ρ indicated all of the results were significant and was less than 0.0001 for every calculation reported here. We have presented results here for the intercept-only model, and will present clustered observation and growth curve models as part of our future work.

Table 1. Table of results from the first core calculus course

Institution	UN(1,1)	SE	Residual	SE	ICC	Intercept	SE
ALL	0.2507	0.008947	1.4429	0.00835	0.148028	2.3047	0.009864
1	0.1903	0.01605	1.5184	0.02306	0.111371	2.5309	0.01841
2	0.1096	0.0252	1.6834	0.04697	0.061127	1.9803	0.02991
3	0.1958	0.01844	1.5044	0.02356	0.115163	2.1321	0.02369
4	0.155	0.01523	1.3719	0.02023	0.101513	2.4528	0.02274
5	0.1695	0.0177	1.3168	0.01593	0.114042	2.5186	0.02334
6	0.3222	0.04311	1.6478	0.04825	0.163553	2.3323	0.04033
7	0.04731	0.008621	1.2952	0.01894	0.03524	2.3486	0.02294
8	0.4001	0.0524	1.7531	0.04348	0.185816	2.0184	0.04819
9	0.1743	0.02918	1.5974	0.04361	0.09838	1.8378	0.03316

Table 2. Table of results from the first core chemistry course

Institution	UN(1,1)	SE	Residual	SE	ICC	Intercept	SE
ALL	0.3899	0.01396	1.2209	0.007642	0.242054	2.4514	0.01313
1	0.2668	0.01654	0.9013	0.01319	0.228405	2.9399	0.01897
2	0.09258	0.0194	1.5112	0.03618	0.057726	2.1721	0.02417
3	0.05687	0.00866	1.2014	0.0162	0.045197	2.3037	0.02168
4	0.08957	0.0187	1.1998	0.02855	0.069468	2.548	0.03098
5	0.2365	0.02995	1.35	0.01858	0.14907	2.3691	0.03912

6	0.5266	0.0616	1.4319	0.0378	0.268879	2.2361	0.04377
7	0.06628	0.01555	0.9488	0.01808	0.065295	2.7636	0.03437
8	0.3878	0.06471	1.6522	0.05703	0.190098	1.8637	0.06203
9	0.1187	0.02301	1.5405	0.03828	0.071541	1.7618	0.03103

Table 3. Table of results from the first core physics course

Institution	UN(1,1)	SE	Residual	SE	ICC	Intercept	SE
ALL	0.2238	0.01143	1.1259	0.007235	0.165815	2.4221	0.01405
1	0.1708	0.0199	0.9418	0.01854	0.153514	2.6287	0.02693
2	0.05667	0.01809	1.211	0.03829	0.044704	2.3403	0.02883
3	0.1616	0.01954	1.307	0.01974	0.110037	2.1877	0.03016
4	0.0827	0.01439	1.2179	0.01954	0.063586	2.3778	0.0327
5	0.1252	0.01692	1.2577	0.01894	0.090534	2.3858	0.02867
6	0.6472	0.08008	1.2458	0.03859	0.341891	2.5402	0.05799
7	0.02418	0.004712	0.8266	0.01096	0.028421	2.7459	0.01902
8	0.3066	0.06185	1.3118	0.03934	0.189446	1.8588	0.07346
9	0.2737	0.1001	0.9619	0.07792	0.221512	2.5246	0.1132

Discussion and Conclusion

Several conclusions can be drawn by examining Tables 1, 2, and 3. These tables show tremendous variation in the magnitude of the ICC between different institutions. These range from as little as 0.028 to as much as 0.34. This corresponds to section accounting for as little as 2.8% of the total variance in final grade to as much as 34% of the variance. There is also variation in the intercept term for each institution, which corresponds to the grand mean of the grades assigned for a course at a given institution. Institutions 2 and 7 are notable in that they both display very low variance across all three of the courses investigated in this study. This could imply that these institutions have instituted policies to reduce sectional differences in their introductory courses via conscious coordination between sections. The coordination of course grades by a single instructor, a single professor teaching all of the sections of a class, or an instructor or instructional team that does not change over multiple years may reduce variance between sections. Not having access to this information is a limitation of this data set. However, these results clearly indicate that the section of a given class can have a significant impact on student achievement, a result consistent with the assertions of Vogt, Felder, and Astin.[6, 8, 30]

In addition to low ICCs, institution 7 also has high intercept values across all three courses. This suggests that institution 7 has a range restriction due to giving out uniformly higher grades and passing more students than comparable institutions.

Grading schemes can also have a significant effect on variance. In normative grading, students are graded in comparison to their peers and grade assignment is generally governed by a desired distribution of final grades. If these distributions are by section such that each section has the same distribution, this would have the effect of minimizing variance between sections. This would also make it nearly impossible to judge the relative teaching abilities of different instructors based on student grades. On the other hand, if the distribution is across all of the sections of a course in a given semester, sectional variance could be much greater with stronger sections assigning a greater proportion of high grades and weaker sections assigning more low grades.

In contrast to normative grading schemes, criterion grading evaluates students based on their mastery of a predetermined set of learning outcomes. When this grading scheme is employed, discrimination between students is deemphasized in favor of trying to achieve mastery. If the majority of students master the required material, this will result in grade range reduction compared to classes employing a normative grading scheme.

It is also worth noting that there is much less variation in the ICCs for Calculus I than are present in both Physics I and Chemistry I due to a variety of possible reasons. Calculus may be somewhat easier to grade consistently than the other two courses due to a clear "right answer" to both homework and exam problems and general lack of open-ended assignments. On the other hand, Chemistry and Physics can have problems that are more difficult to grade consistently across sections, and laboratory courses leave a significant portion of the grade to the discretion of laboratory instructors who may neither teach the corresponding lecture nor grade consistently across sections.

Beyond simply expressing institutional level effects on student performance, Padilla et al. note in their 2005 paper the importance of eliminating aggregation bias and misestimated standard errors that occur when researchers ignore the nested structures inherent in HLM.[36] The treatment of HLM in light of aggregation errors by Padilla is elegantly explained as "when an explanatory variable can take on different meanings." (ibid, p. 2) HLM measures mean values at each level, so while student grade averages within a section have one meaning, their average at the institutional level has another meaning. Padilla et al. noted a 19% variance in grades based solely on institutional differences. Their result complements our work, as we noted sectional variability on the same order (this is particularly expected since Padilla et al. studied an earlier version of the same dataset).

Overall, we are pleased with the results from the preliminary stages of our analysis of three core courses endemic to engineering curricula in our data set. We hope that the introduction

of such a strong variance of student grades at the course section level generates significant discussion in the community. Although multiple factors have already been suggested that can result in sectional differences in grades, the use of additional data from MIDFIELD can shed some light on which factors are of greatest importance. A clarification of the factors that result in sectional differences will be the subject of continued quantitative study and represents an opportunity for qualitative study as well.

References

1. Stevens, R., Daniel M. Amos, Lari Garrison, and Andy Jocuns. *Engineering as Lifestyle and a Meritocracy of Difficulty: Two Pervasive Beliefs Among Engineering Students and Their Possible Effects.* in *Proceedings of the American Society for Engineering Education Annual Conference, Honolulu, Hawaii, June 24-27, 2007*. 2007. Honolulu, Hawaii.
2. Adelman, C., *Women and men of the engineering path: a model for analyses of undergraduate careers*1998, Washington, DC: U.S. Dept. of Education : National Institute for Science Education. xii, 105 p.
3. Suresh, R., *The Relationship between Barrier Courses and Persistence in Engineering.* Journal of college student retention, 2007. **8**(2): p. 215-239.
4. Lambert, A., P. Terenzini, and L. Lattuca, *More than meets the eye: Curricular and programmatic effects on student learning.* Research in Higher Education, 2007. **48**(2): p. 141-168.
5. Volkwein, J.F., *The Assessment Context*, in *Assessing Student Outcomes: Why, Who, What, How?*, J.F. Volkwein, Editor 2009, Jossey-Bass: San Francisco.
6. Vogt, C.M., *Faculty as a Critical Juncture in Student Retention and Performance in Engineering Programs.* Journal of Engineering Education, 2008. **97**(1): p. 27-36.
7. Jewell, R.T., M.A. McPherson, and M.A. Tieslau, *Whose fault is it? Assigning blame for grade inflation in higher education.* Applied Economics, 2011. **45**(9): p. 1185-1200.
8. Felder, R.M. and L.K. Silverman, *Learning and Teaching Styles in Engineering Education.* Engineering Education, 1988. **78**(7): p. 674-681.
9. Wayne, A.J. and P. Youngs, *Teacher Characteristics and Student Achievement Gains: A Review.* Review of Educational Research, 2003. **73**(1): p. 89-122.
10. Deleeuw, J. and I. Kreft, *RANDOM COEFFICIENT MODELS FOR MULTILEVEL ANALYSIS.* Journal of Educational Statistics, 1986. **11**(1): p. 57-85.
11. Cronbach, L.J. and N. Webb, *Between-class and within-class effects in a reported aptitude * treatment interaction: Reanalysis of a study by G. L. Anderson.* Journal of Educational Psychology, 1975. **67**(6): p. 717-724.
12. Luke, D.A., *Multilevel Modeling*. Quantitative Applications in the Social Sciences, ed. M.S. Lewis-Beck2004, Thousand Oaks, California: Sage Publications.
13. Raudenbush, S.W. and A.S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods* 2001, Thousand Oaks, California: Sage Publications.

14. Singer, J.D. and J.B. Willett, *Modeling Change and Event Occurence*2003, New York: Oxford University Press.
15. Raudenbush, S. and A.S. Bryk, *A HIERARCHICAL MODEL FOR STUDYING SCHOOL EFFECTS.* Sociology of education, 1986. **59**(1): p. 1-17.
16. Porter, S.R., *Institutional structures and student engagement.* Research in Higher Education, 2006. **47**(5): p. 521-558.
17. Brix, G., et al., *Pharmacokinetic analysis of tissue microcirculation using nested models: Multimodel inference and parameter identifiability.* Medical Physics, 2009. **36**(7): p. 2923-2933.
18. Cudeck, R. and J.R. Harring, *Analysis of nonlinear patterns of change with random coefficient models*, in *Annual Review of Psychology*2007. p. 615-637.
19. Brostrom, G. and H. Holmberg, *Generalized linear models with clustered data: Fixed and random effects models.* Computational Statistics & Data Analysis, 2011. **55**(12): p. 3123-3134.
20. Dinu, R.R. and A. Veeraragavan, *Random parameter models for accident prediction on two-lane undivided highways in India.* Journal of Safety Research, 2011. **42**(1): p. 39-42.
21. Mohamed, A.S.A. and T. Prasad, *ANALYSIS OF THE NEUROMUSCULAR SYSTEM USING RANDOM PARAMETER MODELS.* Mathematical and Computer Modelling, 1990. **13**(4): p. 19-36.
22. Craine, R. and A. Havenner, *Classical versus Bayesian models-on the dangers of a little bit of knowledge.* International Journal of Systems Science, 1983. **14**(8): p. 871-5.
23. Rodrigues Neto, C. and A.C.R. Martins, *Multifractality in the random parameter model for multivariate time series.* Physica A: Statistical Mechanics and its Applications, 2009. **388**(11): p. 2198-2206.
24. Henderson, C.R., *Estimation of variance and covariance components.* Biometrics, 1953: p. 226-252.
25. Astin, A.W. and J.J. Lee, *How risky are one-shot cross-sectional assessments of undergraduate students?* Research in Higher Education, 2003. **44**(6): p. 657-672.
26. Coleman, J., T. Hoffer, and S. Kilgore, *Cognitive Outcomes in Public and Private Schools.* Sociology of education, 1982. **55**(2/3): p. 65-76.
27. Rogosa, D., D. Brandt, and M. Zimowski, *A growth curve approach to the measurement of change.* Psychological Bulletin, 1982. **92**(3): p. 726-748.
28. Statistics, N.C.f.E., *High School and Beyond: Sample Design Report*, 1980: Washington D.C.
29. Morrell, C.H., L.J. Brant, and L. Ferrucci, *Model Choice Can Obscure Results in Longitudinal Studies.* The Journals of Gerontology, 2009. **64A**(2): p. 215-22.
30. Astin, A.W. and N. Denson, *Multi-Campus Studies of College Impact: Which Statistical Method is Appropriate?* Research in Higher Education, 2009. **50**(4): p. 354-367.
31. Sanders, W.L. and J.C. Rivers, *Research-in-Progress Report.* International Journal of Selection and Assessment, 1996. **4**(2): p. 106-109.
32. Vogt, C.M., D. Hocevar, and L.S. Hagedorn, *A Social Cognitive Construct Validation: Determining Women's and Men's Success in Engineering Programs.* The Journal of higher education, 2007. **78**(3): p. 337-364.
33. Astin, A.W., *What matters in college?: four critical years revisited.*1993, San Francisco: Jossey-Bass
34. Borchers, A. and S. Hee. *Assessing the effectiveness of entrepreneurial education programs from a multi-level multi-dimensional perspective with mental models*. in *118th ASEE Annual Conference and Exposition, June 26, 2011 - June 29, 2011*. 2011. Vancouver, BC, Canada: American Society for Engineering Education.
35. Ohland, M.W., et al., *Persistence, engagement, and migration in engineering programs.* Journal of Engineering Education, 2008. **97**(Compendex): p. 259-278.

36. Padilla, M.A., et al. *Drawing valid inferences from the nested structure of engineering education data: Application of a hierarchical linear model to the SUCCEED longitudinal database*. in *2005 ASEE Annual Conference and Exposition: The Changing Landscape of Engineering and Technology Education in a Global World, June 12, 2005 - June 15, 2005*. 2005. Portland, OR, United states: American Society for Engineering Education.
37. Lawanto, O. and S. Johnson. *Students' cognitive self-appraisal, self-management, and the level of difficulty of an engineering design project: Are they related?* in *2009 ASEE Annual Conference and Exposition, June 14, 2009 - June 17, 2009*. 2009. Austin, TX, United states: American Society for Engineering Education.
38. Harper, B. and P. Terenzini. *The effects of instructors' time in industry on students' co-curricular experiences*. in *2008 ASEE Annual Conference and Exposition, June 22, 2008 - June 24, 2008*. 2008. Pittsburg, PA, United states: American Society for Engineering Education.
39. Strayhorn, T. *Measuring the educational benefits of diversity in STEM education: A multi-institutional survey analysis of women and underrepresented minorities*. in *2009 ASEE Annual Conference and Exposition, June 14, 2009 - June 17, 2009*. 2009. Austin, TX, United states: American Society for Engineering Education.
40. Long, R.A. *The Multiple-Institution Database for Investigating Engineering Longitudinal Development*. 2008; Available from: https://engineering.purdue.edu/MIDFIELD.
41. De Clemencia Cosentino, C. and N. Deterding, *Widening the Net: National Estimates of Gender Disparities in Engineering.* Journal of Engineering Education, 2009. **98**(Compendex): p. 211-226.
42. National Science Board. *Science and Engineering Indicators, 2008*. 2008; Available from: http://www.nsf.gov/statistics/seind08/c0/c0i.htm.