# AC 2009-1022: UNDERSTANDING FACTORS CONTRIBUTING TO RETENTION IN ENGINEERING: A STRUCTURAL EQUATION MODELING (SEM) APPROACH

**Mark Urban-Lurain, Michigan State University**

Mark Urban-Lurain is the Director of Instructional Technology Research & Development in the Division of Science and Mathematics Education at Michigan State University. Dr. Urban-Lurain's research interests are in theories of cognition, their impact on instructional design and applying these to the use of instructional technology. He is also interested in the role of technology in educational improvement and reform.

**Jon Sticklen, Michigan State University**

Jon Sticklen is the Director of the Applied Engineering Sciences major, College of Engineering, Michigan State University. Dr. Sticklen also serves as the College Coordinator for engineering education research, and is an Associate Professor in the Computer Science and Engineering Department, MSU. Dr. Sticklen has lead a laboratory in knowledge-based systems focused on task specific approaches to problem solving. More recently, Dr. Sticklen has pursued engineering education research focused on early engineering; his current research is supported by NSF/DUE and NSF/ CISE.

**Daina Briedis, Michigan State University**

Daina Briedis is an Associate Professor in the Department of Chemical Engineering and Materials Science at Michigan State University. Dr. Briedis has been involved in several areas of education research including student retention, curriculum redesign, and the use of technology in the classroom. She is a co- PI on two NSF grants in the areas of integration of computation in engineering curricula and in developing comprehensive strategies to retain early engineering students. She is active nationally and internationally in engineering accreditation and is a Fellow of ABET.

**Neeraj Buch, Michigan State University**

Dr. Neeraj Buch is a Professor in the Department of Civil and Environmental Engineering at Michigan State University. He is also the Director of Cornerstone Engineering and Residential Experience program at Michigan State University. He earned his M.S. degree in pavement engineering in 1988 from the University of Michigan, Ann Arbor and his Ph.D. in pavement and materials engineering from Texas A&M University, College Station, in 1995. Dr. Buch began his academic career at Michigan State University in 1996. Dr. Buch teaches undergraduate and graduate courses in concrete materials and pavement engineering. He is also involved in teaching short courses on pavement design and rehabilitation and pavement materials for practicing engineers in Michigan. He is a co-PI on two National Science Foundation grants in the areas of integration of computation in engineering curricula and in the area of retention of early engineering students.

**Thomas Wolff, Michigan State University**

Thomas F. Wolff is Associate Professor of Civil Engineering and Associate Dean of Engineering for Undergraduate Studies at Michigan State University. He has taught undergraduate and graduate courses in geotechnical engineering and reliability analysis. His research and consulting has focused on the design and evaluation of dams, levees and hydraulic structures, and he has been involved in several studies related to the failure of New Orleans levees in hurricane Katrina. As Associate Dean, he oversees curriculum, advising, career planning, study abroad, early engineering and other related initiatives.

# Understanding Factors Contributing to Retention in Engineering: A Structural Equation Modeling (SEM) Approach

## Introduction

Retention of early engineering students is a nation-wide concern that will affect the strength of the future engineering workforce and, hence, the role of the United States as a dominant world player in engineering and technology[1]. Michigan State University (MSU) and Lansing Community College (LCC) were recently awarded a five-year NSF STEP grant (**S**TEM **T**alent **E**xpansion **P**rogram) to increase retention by 10% over current levels at our large, research-intensive institution. The project is titled **E**ngaging **E**arly **E**ngineering **S**tudents to Expand Numbers of Degree Recipients (EEES).

The major research challenge in this project is to understand the interactions among the various components of the project. Our engineering curricula are not lock-step, so students may elect to participate in various programs and the interactions among the interventions may vary by student and the choices they make. These challenges make traditional statistical techniques difficult to use.

Structural Equation Modeling (SEM) is a multivariate procedure that supports hypothesis-testing of causal models in observational studies without the need for random assignment of participants to treatment and control groups. This paper outlines our project, introduces SEM and the models we are evaluating and discusses the data collection, management and analysis we are implementing to track the various components of the project. The methods are appropriate for other *in situ* studies of educational interventions.

## Overview of EEES

EEES targets two groups of students who are at-risk for leaving engineering: 1) students who are academically capable of completing an engineering degree but perceive the education environment of early engineering as being unsupportive and not engaging [2-4]; and 2) students who struggle with core prerequisite courses, mainly calculus and physics. Analysis of our past student retention patterns show that grades in these core courses are the best predictors of future admission to Engineering in the junior year. The goal of the EEES project will be achieved through the synergistic deployment of four components designed to involve engineering faculty in rethinking the structure of the introductory courses.

EEES has four components as shown in Figure 1. They are: 1) Connector faculty; 2) Peer Assisted Learning (PAL); 3) Course cross linkages; and (4) Diagnostic (DX) Driven Early Intervention. Details of these components of the EEES project are available in other papers[5, 6].

EEES targets students in four key technical courses taken by early engineering students to prepare them for upper level disciplinary courses: pre-calculus algebra and trigonometry (MTH 116), calculus 1 (MTH 132), physics 1 (PHY 183) and computation-based problem solving (EGR 102). The four EEES components are shown in the bubbles in Figure 1. The dark boxes show groups of faculty or students who are part of the implementation of EEES. The dark bubble shows the target group: early engineering students.

Since the mathematics and physics courses are outside of the College of Engineering, the EGR 102 course is the primary locale for engaging engineering faculty in 1) changing the curriculum; 2) incorporating more active-learning, team-based projects, and integrating mathematics and physics into the curriculum; and 3) fostering a sense of engineering community among faculty and students.

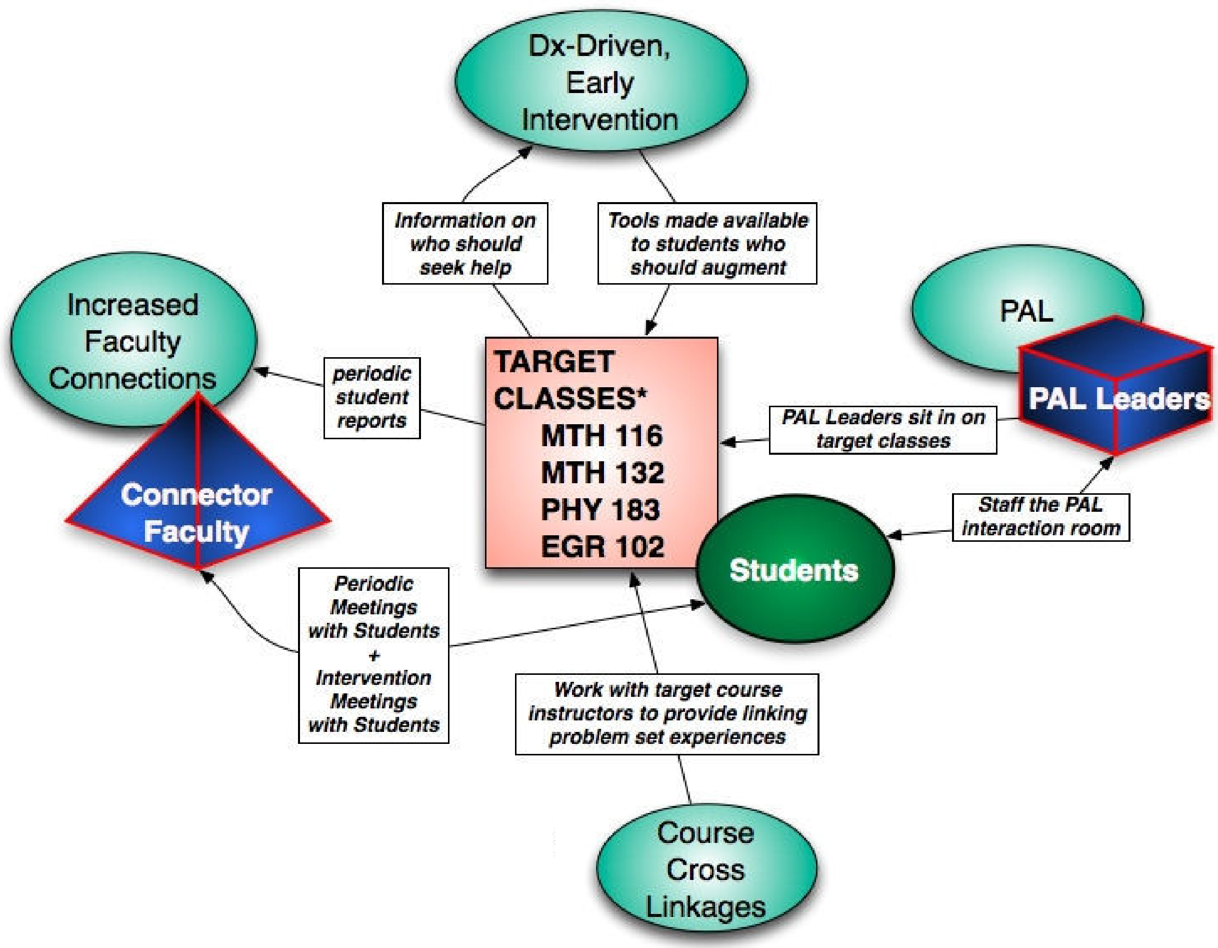


Figure 1: EEES Schematic Diagram.

To support students who are struggling with the crucial calculus courses, we implemented two interventions. The first is a series of online diagnostic (DX) tests that provide formative feedback to students about the areas in which they need to focus their studying. By providing early and frequent opportunities for students to check their progress, the DX program allows students to seek additional help with these courses. To provide that assistance, we implemented a program of Supplemental Instruction [7, 8], which we have called Peer Assisted Learning. The goal of this program is to provide undergraduate peer-led study sessions that target <u>difficult courses</u>, not at-risk students.

In addition to addressing academic issues, some of the interventions we designed for EEES are based on literature that shows that many students leave STEM disciplines not because they are academically incapable, but because they find the academic environment in these disciplines unsupportive and have serious reservations about future careers in STEM disciplines [2-4]. To address these needs, we created a Connector Faculty program that pairs incoming first year engineering students with engineering faculty[5]. The goal of this program is to create a welcoming atmosphere for students where they can feel connected to the community of faculty and upper division students in the college.

**Research Questions**

Our research questions include:

1. What are the changes in retention as a result of EEES?
2. What are the impacts of each of the components of EEES on retention? That is, what proportion of the variance is accounted for by each component?
3. Which groups of students are impacted by which components of EEES? Do different EEES components have different impact on students who are not struggling academically but might leave due to the climate in the College of Engineering compared with students who are struggling academically? Do different components have different impact on women and underrepresented groups?
4. What are the interactions among the EEES components for different groups of students? That is, does participating in multiple components have an additive, multiplicative, or negative interaction? Does this differ by types of students?

Answers to these questions will be crucial to the formative evaluation of EEES, for continuous quality improvement on the program, and to provide a basis for helping students decide what components of EEES to draw upon.

**Research Challenges**

The major research challenge in this project is to understand the interactions among the different types of at-risk students, different interventions, and final outcomes over the course of the project. Since the project is being implemented college-wide, it is not feasible to compare contemporaneous students who are and are not participating in the interventions. Furthermore, students may elect to participate in various parts of the program (or not) and the interactions among the interventions may vary by student and the choices they make. These challenges make traditional statistical techniques difficult to use.

In traditional statistical analysis, when we want to predict future successful admission to engineering based on the grades in calculus and physics, we could create a regression equation. For example, we would define $\mathbf{Y}$ as a binary dependent variable (admit/no admit), and $\mathbf{X_i}$ as the independent variables of grades in the *i-th* calculus and physics courses of interest. Each of the independent variables has a weight, $\boldsymbol{\beta_i}$, that is calculated to reduce the <u>overall</u> error, $\boldsymbol{\varepsilon}$, of the equation. The resulting regression equation would be:

$$\mathbf{Y = \boldsymbol{\beta}_1X_1 + \boldsymbol{\beta}_2X_2 + \ldots + \boldsymbol{\beta}_iX_i + \boldsymbol{\varepsilon}}$$

In regression, the assumption is that the individual measurement errors are randomly distributed and that all of the variance which is unaccounted for is subsumed in a single error term, $\boldsymbol{\varepsilon}$. However, while this equation captures the observed variables ($\mathbf{X_i}$), we have little information that allows us to understand what factors influenced these variables. Furthermore, regression and other multivariate techniques assume that the independent variables are in fact, independent. That is, the values of $\mathbf{X_1}$ and $\mathbf{X_2}$ should not be correlated. Yet, we know from previous studies that students who do well in one course tend to do well in other courses. At our university, overall GPA correlates with grades in individual courses at about $r = .6$ or more, depending on the course.

As noted above, when we analyzed the grades of students who were previously accepted or declined admission to engineering, we found that grades in calculus courses (followed closely by physics) had the strongest correlation with admission to engineering and final GPA at graduation. The grades in these courses are the *observed variables*, which we infer are measuring some *latent variable* such as *mathematical knowledge*. It is the underlying latent variable that is of interest but we have only the proxy variable of a grade in a course that has been measured. As we plan an intervention, we must be careful not to focus only on the observed variable (course grade) but to intervene in a manner that will affect the latent variable (mathematical knowledge) which, in turn, should result in a higher course grade. If we simply focus on improving the course grade, then it is possible that we will help students improve their grades in the target courses, only to face a subsequent hurdle in later courses because they do not have the necessary mathematical knowledge.

## Structural Equation Modeling

Structural Equation Modeling (SEM) is a multivariate procedure that supports hypothesis-testing of causal models in observational studies without the need for random assignment of participants to treatment and control groups[9]. By focusing on covariance structures and including explicit error terms in the models, rather than trying to minimize residuals as in traditional multivariate techniques[10], it is possible to estimate latent variables (e.g., engagement, persistence, academic ability) that are theorized to be the "causes" of observed variables (e.g., grades, retention). Generally SEM is used for hypothesis testing to see how well a hypothesized model fits the data, rather than for exploratory model building. However, it is possible to compare different models as a data exploration technique. SEM is a relatively recent technique (Bollen[10] provides historical background) that gained wider acceptance with the creation of the LISREL software to support the calculations needed. SEM does not appear to have been used very much in Engineering Education research. A search of ASEE conference proceedings for the terms "structural equation modeling" and "SEM" did not turn up any papers since 1996. Vogt[11] recently used SEM to model the impact of faculty interactions with students on retention in engineering.

SEM includes other techniques such as path analysis and factor analysis and usually represents the models by a *path diagram* rather than in the form of equations. SEM path diagrams show the relationships among the variables and can be used to represent standard regression models. Figure 2 shows the path representation of the previous regression equation that we created using AMOS[12], software that supports SEM analysis.

In Figure 2, each of the observed independent variables – grades in the various courses – is shown by a box on the left side of the figure. The dependent variable, *admit*, is show by the box in the middle of the diagram. The proportional variances (r-square) for each of the independent variables are shown by the straight arrows labeled *r2*. The error term is shown by the circle because it is a *latent variable*. In other words, we presume that there is error, but we do not explicitly identify or measure it – it is simply everything not measured by the observed variables plus any measurement error. The straight arrows represent that the independent variables plus the error term account for the variance of the dependent variable. The curved arrows represent the interactions, or covariance, among the independent variables.

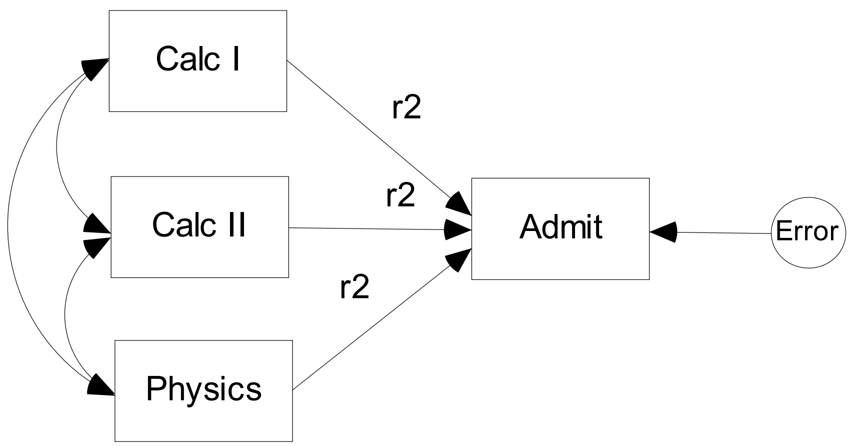


Figure 2: Path representation of regression equation

To analyze data with AMOS, the user specifies a model graphically as shown above, then attaches variables from a dataset to the boxes. The software then solves the model and calculates the beta weights for the independent variables in a traditional least-squares manner.

So far, using SEM in this manner does not provide any particular advantage over traditional statistical techniques. However, the advantage of SEM emerges when we want to analyze more complex situations that are not easily modeled as a single linear equation among observed variables. Building on the previous example, when we use regression to model the admission decision, all of the error is collapsed into a single error term. Often, the predictor variables (i.e., the grades in the courses) can account for the majority of the variance, but the parts of the unaccounted-for variance are all collapsed into the single error term. In practice, there are likely differing amounts of error associated with the different variables. In addition, the grades in the individual courses are inter-related ($r \approx .6$), which can cause some prediction instabilities in the model. One reason for this inter-relationship among the course grades is that the observed variables (grades) are a manifestation of latent (unobservable) variables such as academic ability or motivation. Using SEM, we can re-specify a hypothetical model as shown in Figure 3.

The model shown in Figure 3 is a modification of previous model shown in Figure 2. Here, the interaction terms among the observed variables have been removed and two latent variables are added: *academic ability* and *motivation*. We cannot directly measure these variables, but they are hypothesized to be the causal variables behind the observed variables of grades in each of the courses. These latent traits are assumed to be characteristics of each individual student that affect student outcomes in the course. Academic ability and motivation are shown as co-varying by the double-headed arrow because they are hypothesized to influence each other. That is, students who are motivated, tend to work harder and perform better. But students who perform better receive positive feedback which in turns motivates them. Academic ability and motivation are both shown as affecting course grades, but each one has separate arrow because the weights of these are computed independently. It may be that motivation has differing influence on performance in different courses, so these are allowed to vary independently. By acknowledging that measurement is error-prone and including measurement errors for each of the observed variables, SEM makes it possible to provide unbiased estimates for the latent variables (academic ability and motivation.)

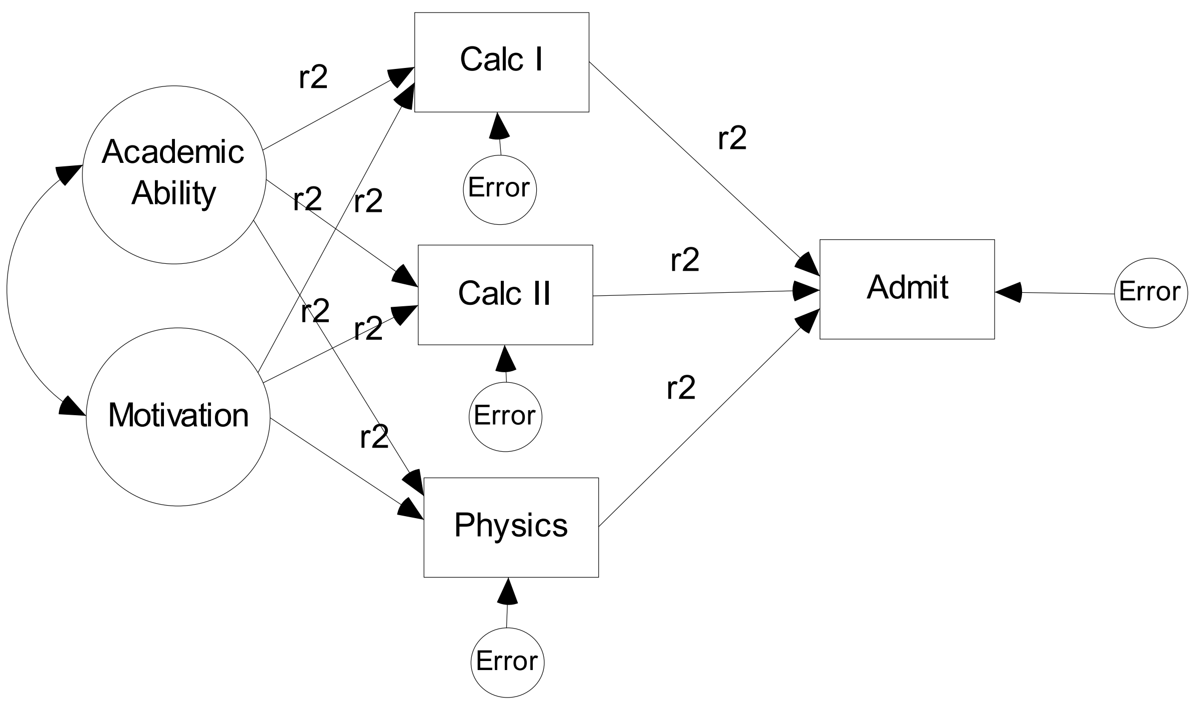


Figure 3: Re-specified model using latent variables

In addition to including the latent variables, SEM also allows the specification of error terms on each observed variable, rather than a single error term for the entire model. This facilitates modeling the affect of the latent variables by partitioning the error across the different independent variables by minimizing each error term in order to calculate the optimal weights for each of the paths in the model.

## A SEM Model for EEES

We are in the first year of the EEES project and are beginning to implement the various components of the project: Connector Faculty, Peer Assisted Learning, DX Intervention, and Cross-Course Linkages. These changes are occurring within the larger setting of ongoing programs throughout the college of engineering. However, all new incoming students are participating in these programs. Therefore, we cannot compare students (or cohorts of students) who are participating in the various components of EEES to understand how the individual components work. Furthermore, EEES was designed to be a series of interconnected components, so it is unreasonable to try to examine any individual component alone. Yet, we will want to understand the components individually to see what works and to make improvements to the individual components. We believe that SEM has the potential to help us understand the contributions of each component plus the interactions among the components.

To address these research needs, we have begun to collect a large amount of data on students and faculty in the College of Engineering. These data include student demographic data (class standing, gender, ethnicity, ACT scores); assignment, exam and course grades in the core pre-engineering courses; course grades in other courses; math diagnostic exam scores; number of PAL sessions attended; number of meetings with Connector Faculty, surveys/interviews on student perceptions of the various components of the EEES program; surveys/interviews on student engagement and desire to persist in engineering. We are tracking students from their first year in the program through one of three outcomes: 1) students who graduate with an engineering degree; 2) students who are academically qualified for engineering but elect to leave for other majors; and 3) students who apply but are not admitted to engineering due to insufficient GPA. Our goal is to understand what, if any, affect the EEES program has on these three types of students and how different students respond to the various components.

Our hypothesized model is shown in Figure 4. Recall that the observed variables are shown in boxes. These are the variables we can measure and include our outcome variables (degree completion/leaving/no admit), other observable variables (demographics, grades, test results, number of PAL sessions and CF meetings attended) and indirect measures of engagement and persistence (surveys and interviews.) Each observed variable has measurement error shown by the black circle with an arrow pointing to it.

Persistence is necessary if a student is going to succeed in any academic discipline, but particularly in one as demanding as engineering [13-16]. The ability to learn from mistakes and change tactics is crucial but requires ongoing support. We anticipate that the support of the peer mentors in the PAL program and interactions with the Connector Faculty will help motivate students to persist in the face of these challenges and help them acquire the skills needed to become successful engineers.

Engagement is crucial to support students' intrinsic motivation, which, in turn, is crucial to learning [13, 17, 18]. While the challenge of problem-solving can be engaging for engineers, supporting new students and helping engage them requires more than the far-off goal of engineering work after graduation. It is important to help new students feel welcomed as part of the community of engineers and scaffold their engagement as part of that community [19]. Both the Connector Faculty and PAL components of EEES are designed to do that.

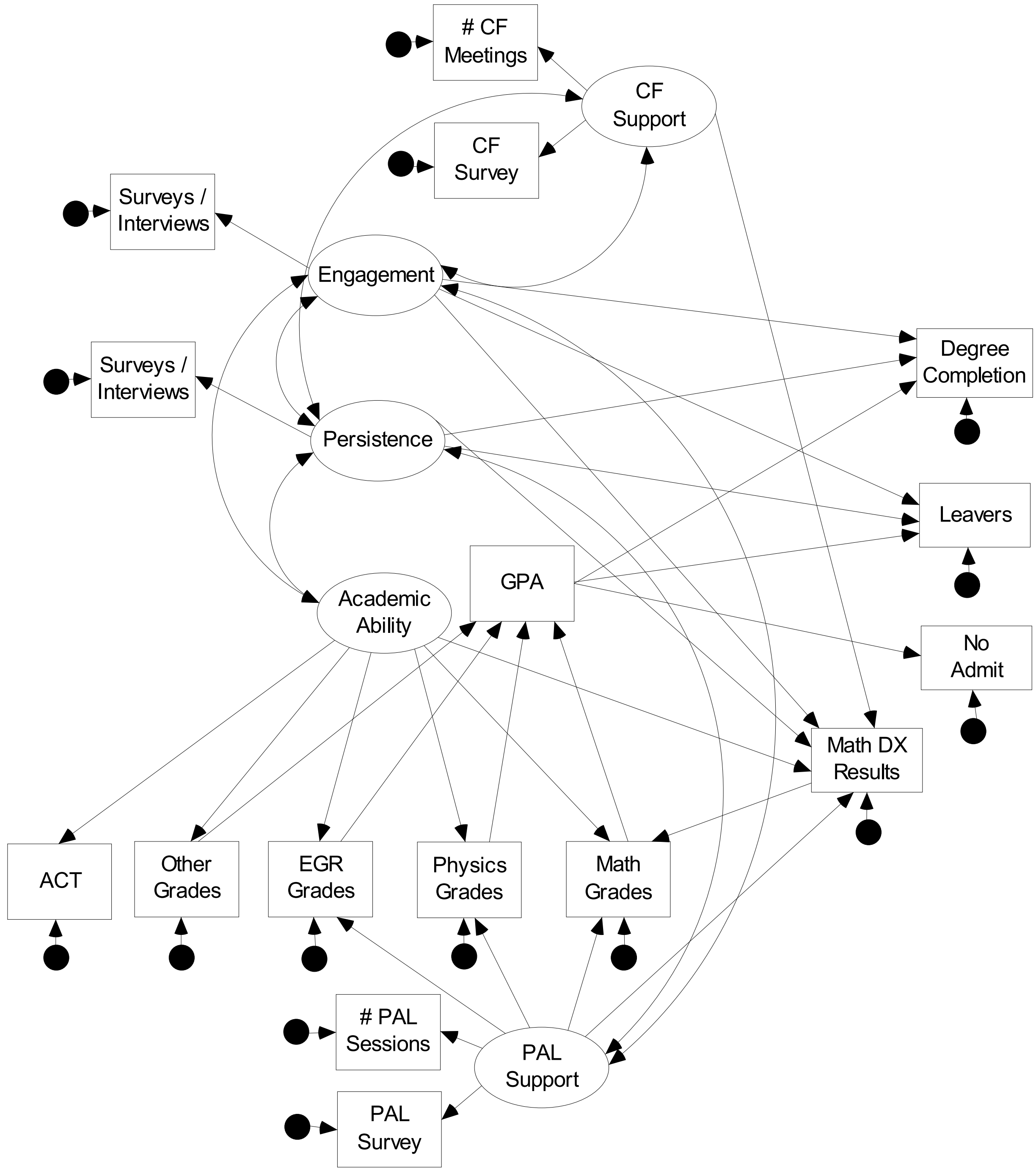


Figure 4: Hypothesized Model for EEES

Academic ability, persistence and engagement are not independent of each other. All three interact, with persistence stimulating engagement and improving academic ability, academic ability producing better grades which helps engagement, and persistence and engagement motivating persistence and improving academic ability. These interactions are shown by the double-headed arrows among these variables. While we have the established measures of academic ability available to us in the form out grade outcome variables, persistence and engagement are much more elusive. The model imputes the values indirectly from outcome variables and from surveys and interviews designed to understand students' persistence and engagement and the affect of the EEES interventions.

## Anticipated Outcomes

SEM can test hypothesized causal models to determine how well the observed data fit the proposed model. However, the large number of nodes in this model requires substantial datasets. At Michigan State University, we have about 1000 incoming first-year students who express interest in engineering and take the pre-engineering curriculum, but they are not formally admitted to engineering until their junior year. These numbers are sufficient for the complexity of the hypothesized model. However, we will have to track these students throughout their programs before we will have outcome data and can completely test the model. Since we are just completing the first year of the project, we have begun collecting data for this cohort. By the end of next year, we will begin to have seen some attrition and will accept or decline students to engineering so we will be able to begin evaluating the model fit at that time.

When SEM models are fit with data, the software tests the fit of the resulting model using chi-square tests. The goal is to have a good fit, that is, <u>not</u> to reject the proposed model, so we look for $p > .05$ (usually substantially higher) to show that the model fits. If the model fits, then the resulting parameter estimates associated with each path in the model indicate how much of the variance of that part of the model is accounted for by each of the subcomponents. For example, in the proposed model, there are three factors that directly influence degree completion: GPA, persistence and engagement. Of these, only GPA is an observed variable. However, GPA is a function of all of the other grades in the courses, which, in turn, are a function of academic ability and PAL support. However, academic ability is influenced by persistence and engagement which are influenced by PAL support and CF support. Each of these paths will have weights that will tell us the proportional contribution of each of the subcomponents, when considered in the context of the whole system. In that way, we will be able to understand the relative contributions of the components of EEES and use that data to improve and refine the various components.

## Conclusion

Improving student retention in engineering programs requires a multi-faceted approach that addresses academic, social and personal issues that vary across an increasingly heterogeneous student population. The EEES project is taking a four-prong approach to this problem: 1) Connector faculty; 2) Peer Assisted Learning (PAL); 3) Course Cross Linkages; and (4) Diagnostic (DX) Driven Early Intervention. These components are inter-related and cannot be evaluated in isolation from each other, presenting a challenge for traditional statistical techniques. In this paper we have presented an introduction to Structural Equation Modeling (SEM) which can be used to analyze complex, inter-related systems consisting of a mixture of observed and latent variables. SEM makes it possible to propose and test causal models using observational data that educational researchers routinely encounter and can be used in a variety of settings to gain deeper insight into the factors affecting educational outcomes.

## Acknowledgement

This material is based upon work supported by the National Science Foundation under award 0757020 (DUE). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

## References

1. National Science Board *Science and Engineering Indicators 2002*; NSB-02-1; National Science Foundation: Arlington, VA, April, 2002.
2. Bernold, L. E.; Spurlin, J. E.; Anson, C. M., Understanding our students: A longitudinal study of success and failure in engineering with implications for increased retention. *Journal of Engineering Education* 2007**,** 96, (3), 263-274.
3. Seymour, E., Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education* 2002**,** 86, (1), 79-105.
4. Seymour, E.; Hewitt, N. M., *Talking about leaving: why undergraduates leave the sciences*. Westview Press: Boulder, Colo., 1997; p x, 429.
5. Briedis, D.; Buch, N.; Collins-Eaglin, J.; Erlich, N.; Fleming, D.; Hinds, T.; Sticklen, J.; Urban-Lurain, M.; Wolff, T. F. *Connector faculty: A friendly face for early engineering students*, ASEE Conference and Exposition, Austin, TX, June 14-17, 2009; IEEE: Austin, TX, 2009.
6. Sticklen, J.; Wolff, T.; Bauer, W.; Briedis, D.; Buch, N.; Courtney, J.; Ehrlich, N.; Fleming, D.; Heckman, R.; Mickelson, R.; Paquette, L.; Urban-Lurain, M.; Weil, C. *Engaging early engineering students (EEES): Background and goals of an NSF STEP project to increase retention*, ASEE Conference and Exposition, Austin, TX, June 14-17, 2009; IEEE: Austin, TX, 2009.
7. Blanc, R. A.; DeBuhr, L. E.; Martin, D. C., Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition. *The Journal of Higher Education* 1983**,** 54, (1), 80-90.
8. Kenney, P. A.; Kallison, J. M., Jr., Research studies on the effectiveness of supplemental instruction in mathematics. *New Directions for Teaching and Learning* 1994**,** 60, 75-82.
9. Byrne, B. M., *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates: Mahwah, N.J., 2001; p xiv, 338.
10. Bollen, K. A., *Structural equations with latent variables*. Wiley: New York, 1989; p xiv, 514.
11. Vogt, C. M., Faculty as a critical juncture in student retention and performance in engineering programs. *Journal of Engineering Education* 2008**,** 97, (1), 27-36.
12. Arbuckle, J. L. *Amos 17.0*, SPSS, Inc.: Crawfordville, FL, 2008.
13. Van Blerkom, M. L., Academic perseverance, class attendance, and performance in the college classroom. In ERIC: 1996; Vol. ED 407618, p 11.
14. Hutchison, M. A.; Follman, D. K.; Sumpter, M.; Bodner, G. M., Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education* 2006**,** 95, (1), 39-47.
15. French, B. F.; Immekus, J. C.; Oakes, W. C., An examination of indicators of engineering students' success and persistence. *Journal of Engineering Education* 2005**,** 94, (4), 419-425.
16. Allen, J.; Robbins, S., Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance. *Research in Higher Education* 2008**,** 49, (1), 62-79.
17. Stipek, D. J., Motivation and instruction. In *Handbook of educational psychology*, Berliner, D. C.; Calfee, R. C., Eds. Macmillan Library Reference USA Simon & Schuster Macmillan ;
Prentice Hall International: New York
London, 1996; pp 85-113.
18. Bransford, J. D.; Brown, A. L.; Cocking, R. R., *How people learn: Brain, mind, experience and school*. National Academy Press: Washington, DC, 1999; p 319.
19. Lave, J.; Wenger, E., *Situated learning: Legitimate peripheral participation*. Cambridge University Press: New York, 1991.