# The Impact of Social Integration on Engineering Students' Persistence, Longitudinal, Interinstitutional Database Analysis.

**Mr. Eric L Huerta-Manzanilla, Universidad Nacional Autonoma de Mexico-UNAM**

Eric L. Huerta is a Fulbright visiting researcher at the Multiple-Institution Database for Investigating Engineering Longitudinal Development Lab at Purdue University. He received a B.S. degree in Industrial Engineering from the Mexico National Autonomous University, a M.S. degree in Quality Engineering from Queretaro University in Mexico, and is a Candidate for the Doctor of Engineering degree at the same University. His research interests in engineering education in expertise acquisition and student's persistence. In the application of engineering he is also working in the use of statistical models for problem solving in industry and for quantitative research of social aspects in the education of engineers. Prior to joining the University of Queretaro, Mr. Huerta spent several years working in manufacturing, leading medium size manufacturing plants of automotive components and industrial goods. Coordinated the startup of two new plants in central Mexico and implemented lean manufacturing and advanced quality systems. Mr. Huerta is member of the American Society for Engineering Education, the American Society for Quality and the American Statistical Association. He is a Certified Six Sigma Black Belt and serves as consultant for medium and small technology based industrial firms.

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

Matthew W. Ohland is 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 $12.4 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 2011. Ohland is Past Chair of ASEE's Educational Research and Methods division and a member of the Board of Governors of the IEEE Education Society. He was the 2002-2006 President of Tau Beta Pi.

**Mr. Russell Andrew Long, Purdue University, West Lafayette**

# The Impact of Social Integration on Engineering Students' Persistence, Analysis of a Longitudinal and Multi-institution Database.


## Abstract

The main models used to study persistence are Astin's Theory of Involvement in Higher Education, Pascarella's General Model for Assessing Change and Tinto's Theory of Student Departure. Tinto proposed that academic and social integration reinforce students' commitment to their institution and educational goals improving retention. This claim was assessed with an algorithm for mutuality that evaluates social integration in a network.

It is not known if standard academic records may be used for sociometric techniques applied in engineering education research. This paper will introduce this approach and, in particular, discuss the social network parameter "mutuality" and study its relationship to persistence in engineering. Mutuality is an index that assesses the tendency for individuals in a group to reciprocate choices more frequently than would occur by chance, thus mutuality reflects reciprocity beyond random grouping, due to students having free selection of groups.

The records of the Multiple-Institution Database for Investigating Engineering Longitudinal Development were grouped to establish which students took classes in each other's presence, a simplified mutuality algorithm was evaluated with this data and probability Weibull models were fitted for persisters and non persisters. The models for persisters shown larger mutuality scales with lower shapes than those for non-persisters, meaning that they paired with classmates more frequently than the students leaving.

Results suggest that indexes for social networks may be calculated using standard academic records, facilitating the assessment of social integration and assisting the analysis of the departure puzzle's factors and informing policy making processes in education.

## Introduction

Educators of engineers are facing the declining interest of potential students for the field [1], a lack of diversity of those who study engineering [2], and the need to assure that programs effectively prepare the graduates for the current engineering challenges [3],[4]. These conditions motivate educators to be interested in the understanding of outcomes of engineering programs, and in particular, persistence of engineering students and its relations with factors that can be modified to improve it [5].

The persistence of students in American Colleges has been a research subject for more than fifty years, and its study can be traced back to the work on dropouts of Summerskill [6]. There are theories that have partially explained the causes for attrition. However, it remains a research subject because there is no one definitive comprehensive theory yet. Examples of some of the major theories on attrition are Tinto's Theory of Student Departure [7–9], Astin's Theory of Involvement in Higher Education [10] and Pascarella's General Model for Assessing Change [11],[12]. We have elaborated on Tinto's theory, in particular, through a novel way to operationalize aspects of social and academic integration.

Academic and social integration affect the evolution of subsequent commitment of the student to the institution, and therefore, to the goal of graduation, and thus, to the student's intention and commitment to persist. Tinto's theory considers factors such as the student's pre-college characteristics, playing a role in the college departure process. These characteristics are grouped as family background factors, skills and abilities, and prior schooling. These entry characteristics influence the student's initial commitment to the institution. The institution's standards and structure define the student's academic integration into the college's formal system. Thus, student's entry characteristics, and institution's structure affect the normative, academic and social integration of the student at the institution, when the individual's intellectual development is congruent with the environment of the college [8].

Social integration could be understood as the degree of congruency between the student's social behavior and the social system of the university. Academic, or structural, integration is the academic performance and achievement of the student. Normative integration, according to Tinto, reflects the student's appraisal of the academic system of the university and is evident as part of the student's intellectual development. The goal of college graduation and the commitment to the institution are, in Tinto's model, direct and proportionally dependent on the academic and social integration of the student [13].

Tinto stated that the classroom is at the center of the academic experience, and could be understood as the place and time where and when pivotal activities of the social and academic life happen [9]. The activities that students have in the classroom are still a central part of the structure of the modern college, and may even be the only place and time for building relationships for those students that do not live on campus [14].

This analysis is based on the hypothesis that the intention to meet with particular others at the classroom, and the frequency with which these meetings occur, may reflect higher levels of social integration and, therefore, be related with better persistence levels. This suggest that social integration, which is important to the goal and intention to graduate as predicted by the theories of Tinto, Astin and Pascarella, might be operationalized using the social network concept of mutuality. Mutuality, or reciprocity, is an index that assesses the tendency for individuals in a group to reciprocate choices more frequently than would occur by chance [15]. Mutuality is one of many structural characteristics of a social network that reflects cohesiveness of a group; thus, that is an indicator of social integration.

It is proposed that the sociometric techniques of social network analysis [16] can be adapted for the assessment of the integration of the students in the institutional environment. This paper will introduce this approach and, in particular, proposes an adaptation of the parameter "mutuality" and analyzes its relationship to persistence in engineering.

**The application of social network theory to class enrollment**

The structural analysis of social network data sets can be executed either by translating the theoretic statements about the network into sets of graphs or by the statistical analysis of the stochastic assumptions about relational data contained in the social network dataset. Also, the analysis can be local or global; the first at the graph level, and the second at the entire network level [17]. The work presented here will use the statistical analysis approach, with a probabilistic algorithm to evaluate mutuality at the global level, considering institutions as independent networks and the final level of aggregation.

The dyad, or pair, is the basic structural element of a social network. One of the indices that describes the relationship of dyad is mutuality, or reciprocity. The term dyad is understood as two individuals who can interact, because they are part of a group. An adaptation of the algorithm proposed by Katz and Powell [15] and the idea of standardization principle proposed by Rao and Bandyopadhyay [18] were applied to student records of class enrollment rather than traditional social network data. This strategy aims to establish the feasibility of social network analysis using large existing databases rather than relying on data that is much more difficult to gather.

Katz and Powell proposed estimations for reciprocity and for the expected value – probability – of mutual reciprocal selections between two actors in a network. Equation 1 presents the estimation for the probability of reciprocity [15].

$$t_a = \frac{2(N-1)m - Nd^2}{Nd(N-1-d]} \quad (1)$$

In Equation 1, $N$ is the number of individuals in the group, $d$ is the number of choices the $N$ individuals have, expressed as a fixed number, and $m$ is the count or frequency of reciprocity. The variable $m$ is normally obtained with surveys or through ethnographic methods. For this study, $m$ was obtained counting instances of pairs in the database records. The expected value of reciprocal choices under these assumptions is shown in Equation 2. This expected value is also interpreted as the group having no intentional mutual choices, or the base random reciprocity in the group.

$$E(m) = \frac{Nd^2}{2(N-1)} \quad (2)$$

Thus, $t_a = 0$ means reciprocation only occurred by chance, because the expected value of reciprocity is part of the divisor in the quotient. The estimator $t_a$ is the evaluation of the strength of the tendency of reciprocating choices among a dyad or pair of individuals in a network, if the expected probability of the mutual choice between a dyad is considered fixed. Fixed choice means that the number of reciprocal selections per dyad is restricted to a single response, either positive or negative. This assumption simplifies the estimation, and it was made in this study. The estimator $t_a$ will range from minus one to one. In set notation, $t_a \in (-1,1)$. When $t_a = 0$ the estimator means that the group shows only random reciprocity, meaning that no intentional selection has been made. Values greater than zero, with maximum at one, mean the reciprocity of the group is intentional. Values below zero, with minimum at minus one, indicate intentional avoidance of reciprocity.

**The research questions**

Two preliminary research topics were elaborated for this work:

1) Is it possible to use academic records to estimate social network indexes?
2) Are these indexes related to persistence?

Records in MIDFIELD do not have sociometric data as they are commonly considered. We propose that using the probabilistic approach to evaluate the structure of a social network allows the usage of academic enrollment information to estimate social network indexes. This paper presents the first such operationalization, evaluating the social index of mutuality. A discussion of the results disaggregated by institution is presented to evaluate if a trend emerges when compared with their persistence information.

The two research questions posed can be integrated in a single inquiry goal as follow: Is it possible to assess the relationship of social integration and persistence by estimating indexes like mutuality, using only academic records?

The justification for such area of inquiry is that schools normally keep complete academic records. Thus, if such information can be used for evaluating an important aspect of academic development, like student's integration, it may allow an interesting use of those hard-built datasets, for institutional strategic analysis, and for policy making and planning purposes.

## Methods

The approach to answer the research question was a quantitative analysis using a database with student and course records. The variables required were a longitudinal tracking identifier for the students in the database and a detailed list of courses attended by each student where a course is identified both more generally by its content area and level as well as more specifically in terms of a particular instance when an offering is available for enrollment. This combined information is referred to as a group-class. The first stage of the work was the preparation of a dataset by selecting the appropriate variables from a much larger dataset, and then by filtering the records to only keep engineering students records that met conditions that support the reasoning for the probabilistic algorithm for mutuality evaluation. A second stage was the implementation of the algorithm and the estimation of aggregated empirical probabilistic cumulative distribution function of mutuality for each institution.

Rao and Bandyopadhyay [18] proposed that reciprocity is a measure of stability of a social network, and a measure of interdependence between pairs in it . An emphasis of Rao and Bandyopadhyay is the requirement for the index to be standardized to make it of comparable magnitude across different networks. Thus in our work, we standardized the results, and the final aggregated results were reported as empirical probability density functions for each institution, in the range from zero to one, as a standard probability.

Specifically, records from The Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) were used for this study. MIDFIELD is a large database of student records [19]. At the time of the study, the database had records from 1988 through 2011, and held academic information for 978,218 students of eleven public universities in the United States. MIDFIELD does not include indicators for the construction of a traditional sociometric data set. However, based mainly in the definition of mutuality, a pseudo-sociometric evaluation was estimated by counting the frequency of dyads per group-class, as will be explained in detail as the algorithm is described.

### *Stage one: preparing and filtering the dataset*

Records from the MIDFIELD were used for this study. To protect the participating institutions and students, the dataset contains no personally identifiable information. Although both students of each dyadic pair must be from the same institution, all reports are disaggregated while masking institutional identity. The data set used in the study was current up to July, 2012.

The original database had the records for 978,218 students from eleven U.S. universities. The records for eight of these universities were used for the analysis reported in this paper. Three of the participating universities did not provide a variable required for the identification of group-class, the identifier needed to calculate the reciprocity algorithm, so those institutions were omitted from the study. The institutions considered in the analysis do not engage in large-scale clustering of students to build cohorts. Therefore, with the exception of smaller-scale interventions that operate in cohorts (which are unknown and neglected in this study), students have free choice to enroll in any section of a class in their curriculum.

The student records from the eight universities considered were filtered to include only those students with a GPA greater than 2.5, non-transfer students matriculated as engineers, which have never opted for a non-STEM major, and have completed their programs. Students with GPA below 2.5 are more likely to leave their majors because of academic issues [20], so only students with grades above that threshold were considered in the analysis to better isolate the effect of mutuality. Transfer students are likely to present social integration patterns that differ from students who matriculate as first-time-in-college students [21], so they are a subject for future consideration. Those students who at some point opted for a non-STEM major may differ in their profiles from those that were always in an STEM major [8]. Table 1 shows the total number of records in the database for all the universities and for the eight included in the study. The table also shows how many records were from first-time engineering students. The last column shows how many students met the criteria applied to filter the records.

| *Institutions* | *All Students* | *First-Time Engineering Students* | *Students Meeting Criteria* |
|---|---|---|---|
| *11* | 978,218 | 174,082 | 88,624 |
| *8* | 702,532 | 108,145 | 55,995 |

Table 1**.** Students records included in the study and how they relate with original data.

The considerations applied to filter the records reduce the possibility that these other effects will mask the primary effect. A side benefit of the filtering was the reduction of the number of students, which reduced the computational requirements of the proximity algorithms. The subset of students whose records met the filters explained above was 55,995 students of eight universities.

The students' records were joined with the corresponding courses records to get the table with group-class information to count the instances of dyads. Samples of ten percent of the students per class section were selected, and the frequency of their dyads, with students in their class section, was counted. This is a known limitation of the study. The reason for this restriction and the concomitant limitation is mainly the computational intensity of the algorithm. The process of counting dyads generates tables that expand the size of the files in RAM exponentially. Therefore, an analysis that includes all the students in the dataset is left as future work.

***Stage two, first part: The proximity algorithm used for the analysis***

MIDFIELD records do not have information that depicts the social relations of the students. Thus, the analysis was based on probabilistic measures. The probability was calculated based on the frequency of dyads. A dyad was defined as two students who were in the same class section. A dyad with frequency of one was count when two students were in the same class section one time, a count of two out of the same dyad was reported if the same pair of students meets again in a different class section and so on. The count of dyads was accumulated per group-class.

The algorithm that was implemented in this study is presented in Equation 3. It was partially based on Katz and Powell's algorithm. The variable $N$, in the Equation 3, was the number of individuals in the cohort $j$, in a particular academic period, in a particular institution. The variable $m$ was the count of dyads of the cohort $j$, in all the group-class items of the same institution. The $m$ was counted separately for persisters, and non-persisters to analyze separated probability density functions for each type [5]. The variable $d$ was defined as only one possible selection per group-class, for each individual out of the $N$-$1$ persons in a cohort. Thus, the formula to calculate the probability for the mutuality of the cohort $j$ was the Equation 3.

$$t_j = \frac{2m(N-1)}{N^2} \tag{3}$$

Therefore, the discrete probability function $F$ was calculated with the expression in Equation 4 for the $j$ items of the type group-class in the data set, being $k$ the number of cohorts counted for the group-class. The frequency values were disaggregated by institution for this analysis.

$$F_i = \sum_{j=1}^{k} t_j / k \tag{4}$$

The density function was the expression shown in the Equation 5. Where $z$ is the total number of group-class items per institution. The result can be disaggregated by other categorical variables available in the original records of MIDFIELD, such as discipline, class year, gender, and age, with appropriate adjustments in the preparation of the initial data set. Each case should be processed independently and the appropriate qualitative considerations should be applied and justified, for the results to be congruent with the actual conditions of each institution, discipline, class, gender and the like.

$$DPF = \sum_{i=1}^{z} F_i = \sum_{i=1}^{z} \sum_{j=1}^{k} t_j / k, \forall\, j \in \{\mathbf{Z} > 0\} \tag{5}$$

***Stage two, second part: Computational aspects in the analysis***

The MIDFIELD data set was kept in two flat files in the SAS7bDat format. The version of the files used for this analysis was dated July 2012. One of those files had the student information and the other had the course data. The files were 3.1 GB and 2.3 GB in size respectively. The filters described earlier were applied to the files, using SAS®. Smaller files, with only the fields and records required for the analysis, were saved in CSV format. These smaller files were all under 1 GB in size. The main part of the algorithm was implemented with R [22]. The analysis was run for each school in the filtered database, thus, the files were loaded into R querying only the necessary variables for each school run, using the sqldf-package {sqldf} [23]. Therefore, only the appropriate records per school were loaded in RAM. This procedure allowed the use of two files of raw data per run with size under 0.5 GB each, and four more files that were generated by the process. These files were processed with no memory allocation problems in a PC with only 8 GB of RAM, a dual core processor, using a 64 bit operating system, and 64 bit libraries in R. The computing time was under one hour per institutional data set. The code used included data preparation and data processing. A summary table with the frequency data for $m$ was prepared;

probability density functions were estimated and charted. The RAM reported in use by the operating system was under 50% for most of the processing time.

## Findings

The probability density functions obtained are shown in Figure 1. For all but one institution, there is a visible difference between the persisters' and non-persisters' densities. This finding looks potentially interesting, particularly if there is qualitative support for the anomaly at that particular college, because it is unique among the eight institutions in the dataset. The densities for persisters have a wider range in the domain for that institution.

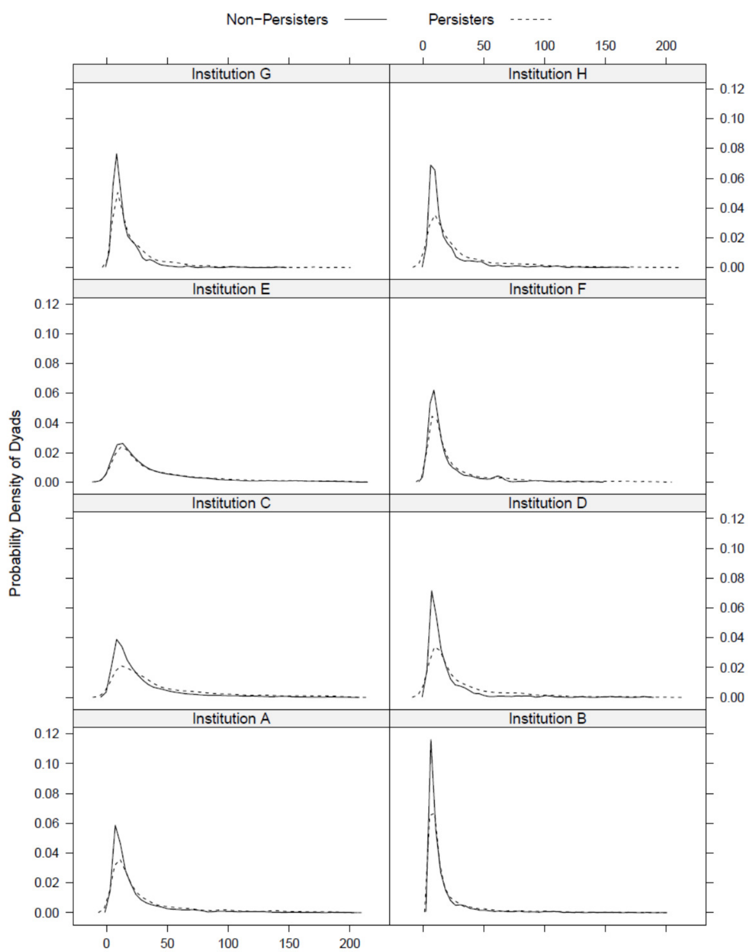


Figure 1. Estimated probability density for mutuality (reciprocity) per institution. All institutions show a difference in densities for persisters and non-persisters, except Institution E.

All the non-persister's densities concentrated around lower values, but persisters appeared to present slightly larger reciprocity values, for most of the density area.

The reciprocity data for each institution was fitted to Weibull probability density functions using the maximum likelihood method. This made it possible to quantify the observations made in the chart. The results are shown in Table 2.

| | **Persisters** | | **Non-Persisters** | |
|---|---|---|---|---|
| Institution | Scale (*m*) | Shape | Scale (*m*) | Shape |
| A (53%) | 15.99 | 0.46 | 7.73 | 0.59 |
| B (38%) | 6.43 | 0.74 | 3.47 | 0.81 |
| C (50%) | 41.67 | 0.58 | 14.29 | 0.65 |
| D (31%) | 13.26 | 0.68 | 4.50 | 0.77 |
| E (57%) | 37.74 | 0.55 | 27.28 | 0.57 |
| F (62%) | 10.32 | 0.69 | 5.09 | 0.77 |
| G (34%) | 9.14 | 0.76 | 4.33 | 0.86 |
| H (26%) | 10.47 | 0.66 | 5.09 | 0.75 |

Table 2. Scale and shape parameters for the Weibull probability distribution fitted for the results of mutuality.

The parameters of the fitted Weibull distributions show quantitative confirmation of the visual differences found in the charts. Shape parameters are greater for non-persisters, indicating that persisters have more variation of reciprocal choices. The distributions for non-persisters have lower dispersion and are more concentrated near their scale parameter. Scale parameters are greater for persisters, indicating that expected values for intentional reciprocity are greater for persisters. For asymmetric distributions, like these, neither mean nor median is of interest; the scale parameter may be understood as the most representative value of mutuality for each institution.

These observed differences were evaluated using correlation and analysis of variance. No correlation was found for scale and shape with the persistence percentages of the engineering students, shown in the column one of Table 2. The analysis of variance for the models shown in Equation 6, have no significance, nor correlation. The results are in Tables 3 and 4. In the second model without interaction the scale factor had $p\text{-}val<0.1$. It is shown in Table 4.

$$persistence \approx \begin{cases} scale + shape + \text{interaction} + \varepsilon; \quad p = 0.245, \quad R_m^2 = 0.283 \\ scale + shape + \varepsilon; \quad p = 0.273, \quad R_m^2 = 0.126 \end{cases} \tag{6}$$

| | *d.f.* | *S.S.* | *M.S.* | *F* | *p-val* |
|---|---|---|---|---|---|
| Scale | 1 | 0.046895 | 0.046895 | 3.1691 | 0.1004 |
| Shape | 1 | 0.020820 | 0.020820 | 1.4070 | 0.2585 |
| Interaction | 1 | 0.002489 | 0.002489 | 0.1682 | 0.6889 |
| Residuals | 12 | 0.177571 | 0.014798 | | |

Table 3 Analysis of Variance for persistence, with scale, shape, and their interaction as factors.

| | *d.f.* | *S.S.* | *M.S.* | *F* | *p-val* |
|---|---|---|---|---|---|
| Scale | 1 | 0.046895 | 0.046895 | 3.3857 | 0.0887 |
| Shape | 1 | 0.020820 | 0.020820 | 1.5032 | 0.2419 |
| Residuals | 13 | 0.180060 | 0.013851 | | |

Table 4 Analysis of Variance for persistence, scale and shape as factors, with no interaction.

The results suggest that persisters and non-persisters may have different reciprocity frequencies for dyads, but only the scale of the probability density functions was found to be significant. The estimated mutuality density functions and their fitted Weibull functions were found to be uncorrelated with the persistence percentages for engineering students.

**Results and discussion**

The findings of this study encourage further development in the application of sociometric analysis for the exploration of academic outcomes. Mutuality as operationalized in this work appears to have some validity since it seems to measure some common factor related to persistence and because it builds upon recognized theories. If mutuality can be calculated from existing institutional course records, social network parameters become much more accessible to institutions.

While some of the limitations described earlier can be addressed in future work, it is important to acknowledge them when interpreting the results of this study. The academic records used for the analysis do not have actual sociometric information and this was the major constraint for the study proposed – we are inferring that intentional mutuality is evidence of social integration.

This fact limits the extension of the arguments that can be inferred from the results. Also, it separates this study and its results from traditional sociometric analysis. It cannot be stated that the reported "selections" were intentional, due to the lack of actual social preferences in the dataset. What we do know is that the students actually were in the same place during most of the meetings for a particular course in a term. A second important limitation was that the analysis was prepared by sampling the database. We know that it is highly desirable to use all the records available, and that larger sample size might result in statistical significance, particularly since there appears to be a qualitative difference in the probability density distributions. Unfortunately, there was no efficient data structure and corresponding algorithm to convert the academic records available into relational data.

The mutuality or reciprocity construct in this study is based on the probability for students that meet during one term to have some degree of interaction. Thus, if they meet a second time, the probability for the students to have met by chance reduced, and so on. Following this line of reasoning, the values of dyads in the same class were counted as intentional only after the fifth. This criteria is independent and in addition to the expected mutuality, or base mutuality, in the Equation 2. This restriction is certainly conservative, and may be so conservative as to have prevented findings of interest.

**Acknowledgement**

This material is based upon work supported by the National Science Foundation under Grant No. 0935157, by The United States - Mexico Commission for the Educational and Cultural Exchange Grant No. 68120602 and by Mexico's National Council for Science and Technology Scholarship No. 293125. 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, nor The United States - Mexico Commission for the Educational and Cultural Exchange, neither those of Mexico's National Council for Science and Technology.

## References

1 Melsa, JL (2007) ‘The winds of change’. *ASEE Banquet Keynote Speech*.

2 Chubin, Daryl E., May, Gary S. and Babco, Eleanor L. (2005) ‘Diversifying the engineering workforce’. *Journal of Engineering Education*, 94(1), pp. 73–86.

3 Lattuca, L. R (2010) ‘INTERDISCIPLINARY COMPETENCE: ALearning GOAL FOR STUDENTS–AND FACULTY?’

4 Litzinger, T., Lattuca, L. R, Hadgraft, R. and Newstetter, W. (2011) ‘Engineering education and the development of expertise’. *Journal of Engineering Education*, 100(1), pp. 123–150.

5 Ohland, M. W, Sheppard, S. D, Lichtenstein, G., Eris, O., et al. (2008) ‘Persistence, Engagement, and Migration in Engineering Programs’.

6 Summerskill, John (1962) ‘Dropouts from College’, in *The American college: A psychological and social interpretation of the higher learning*, Hoboken, NJ, US, John Wiley & Sons Inc, pp. 627–657.

7 Tinto, V. (1987) *Leaving college: Rethinking the causes and cures of student attrition.*, ERIC. [online] Available from: http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED283416 (Accessed 7 January 2013)

8 Tinto, V. (1988) ‘Stages of Student Departure: Reflections on the Longitudinal Character of Student Leaving’. *The Journal of Higher Education*, pp. 438–455.

9 Tinto, V. (2012) *Completing college: Rethinking institutional action*, University of Chicago Press. [online] Available from: http://books.google.com/books?hl=en&lr=&id=zMEy9V4BqDAC&oi=fnd&pg=PR5&dq=theory+of+students+departure+tinto&ots=K47Ya4i0Mi&sig=2s5p6icTpB1ZQ84rC6-laFQR1x8 (Accessed 7 January 2013)

10 Astin, Alexander W (1984) ‘Student Involvement : A Developmental Theory for Higher Education’. *Journal of College Student Development*, (July), pp. 518–529.

11 Pascarella, Ernest T. (1985) ‘Students’ Affective Development within the College Environment’. *The Journal of Higher Education*, 56(6), pp. 640–663.

12 Pascarella, Ernest T. (2006) ‘How College Affects Students: Ten Directions for Future Research’. *Journal of College Student Development*, 47(5), pp. 508–520.

13 Tinto, V. (1975) ‘Dropout from higher education: A theoretical synthesis of recent research’. *Review of educational research*, 45(1), pp. 89–125.

14 Tinto, V. (2000) ‘Linking learning and leaving’. *Reworking the student departure puzzle*, pp. 81–94.

15 Katz, Leo and Powell, James H. (1955) ‘Measurement of the Tendency Toward Reciprocation of Choice’. *Sociometry*, 18(4), pp. 403–409.

16 Scott, John (1988) ‘Social network analysis’. *Sociology*, 22(1), pp. 109–127.

17 Wasserman, S. (1994) *Social Network Analysis: Methods and Applications*, Cambridge University Press.

18 Rao, A. Ramachandra and Bandyopadhyay, Suraj (1987) ‘Measures of Reciprocity in a Social Network’. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 49(2), pp. 141–188.

19 Ohland, M. W, Zhang, G., Thorndyke, B. and Anderson, T. J (2004) ‘The Creation of the Multiple-Institution Database for Investigating Engineering Longitudinal Development (midfield)’. *Proc. Amer. Soc. Eng. Ed., Salt Lake City, Utah*.

20 Bank, B. J., Biddle, B. J. and Slavings, R. L. (1994) ‘First-Semester Grades, Thought Modes, and Undergraduate Persistence’. *Contemporary Educational Psychology*, 19(4), pp. 416–429.

21 Chapman, D. W. and Pascarella, E. T. (1983) ‘Predictors of academic and social integration of college students’. *Research in Higher Education*, 19(3), pp. 295–322.

22 Team, R. Core (2012) *R: A Language and Environment for Statistical Computing*, Vienna, Austria. [online] Available from: http://www.R-project.org/

23 Grothendieck, G. (2012) *sqldf: Perform SQL Selects on R Data Frames*, [online] Available from: http://CRAN.R-project.org/package=sqldf