Model Eliciting Activities motivated problem solving process: solution path analysis

Dr. Natasa S. Vidic, University of Pittsburgh

NATASA VIDIC University of Pittsburgh / Assistant Professor / Department of Industrial Engineering 1032 Benedum Hall / Pittsburgh, PA 15261 412-624-9846 (ofc) / 412-624-9831 (fax) / nvidic@pitt.edu

Professional Preparation University of Pittsburgh Industrial Engineering Ph.D. May 2008 University of Delaware Operations Research M.S. May 1992 University of Belgrade, Serbia Civil/Transportation Engineering B.S. June 1987

Appointments University of Pittsburgh, Pittsburgh, PA Assistant Professor, Department of Industrial Engineering Sept 2010 - present Center Associate Director of Operations, Engineering Education Research Center Jan 2011- Sept 2013 Visiting Assistant Professor Sept 2008 – Sept 2011 Graduate Research Assistant Sept 2002 – Sept 2008 Port Authority of Allegheny County, Pittsburgh, PA Project Manager/Planning Department Aug 1992- Oct 1994 University of Delaware, Delaware Transportation Center, Newark, DE Graduate Research Assistant Aug 1990 - May 1992 University of Novi Sad, Institute for Traffic and Transportation Eng., Serbia Research Associate /Lecturer Aug 1987 – Aug 1990

Selected Publications ● Bursic K., Vidic N., Yildrim T. P., Besterfield-Sacre M., Shuman L., (2013), "The Impact of Simulating Real Life Experiences in Engineering Education: A Case for Model Eliciting Activities", to be submitted to Journal of Engineering Education. ● Vidic, N. and Norman, B. (2013), "Dynamic Assignment: One-cycle and two-cycle math formulations" submitted to International Journal of Operations and Production Management. ● Barbot, E., Vidic, N., Gregory, K., Vidic, R.D. (2013) "Spatial and Temporal Correlation of Water Quality Parameters of Produced Waters from Devonian-age Shale following Hydraulic Fracturing", Environmental Science & Technology, 47, 2562-2569. ● Kelmendi-Doko A., K. G. Marra, N. Vidic, H. Tan, J. P Rubin (2013) "Adipogenic Growth Factor-Loaded Microspheres Increased Retention of Transplanted Adipose Tissue ", submitted to Tissue Engineering. ● Siewiorek N., Shuman L. J., Besterfield-Sacre M., Vidic N., Bursic K. M., (2012). "Students' Confidence Levels in Technical Concept Knowledge with Model Eliciting Activities", Proceeding of the 2012 American Society for Engineering Education Annual Conference, San Antonio, TX. ● Shuman L. J., Besterfield-Sacre M., Bursic K. M., Vidic N., T.P. Yildirim, and N. Siewiorek (2012). "CCLI: Model Eliciting Activities", Proceeding of the 2012 American Society for Engineering Education Annual Conference, San Antonio, TX. ● Vidic N., Shuman L. J., Besterfield-Sacre M., Bursic K. M., T.P. Yildirim, and N. Siewiorek (2011). "Learning Impacts Gained from Introducing Model Eliciting Activities (MEAs) in an Introductory Statistics Course", Proceeding of the Industrial Engineering Research Conference, Reno, Nevada, May 21-25, 2011. ● Larry Shuman, Mary Besterfield-Sacre, Karen Bursic, Natasa Vidic, Pinar Yildirim (2011), "CCLI: Model Eliciting Activities: Experiments and Mixed Methods to Assess Student Learning – Part II")," 2011 Proceeding of the American Society for Engineering Education Annual Conference, Vancouver, B.C. ● Vidic, N. and Norman, B. (2008), "Dynamic Assignment: One-cycle and two-cycle math formulations", INFORMS Conference 2008 (Washington, DC) ● Vidic, N. and Norman, B. (2006), "Dynamics of two-worker two-stations line with worksharing", INFORMS Conference 2006 (Pittsburgh) ● Vidic, N., Norman, B., Besterfield-Sacre, M., Halverson, E. "Understanding and evaluating scheduling practices in urban high schools ", INFORMS Conference 2006 (Pittsburgh)

Research Activities Currently participating in the collaborative research: "Improving Engineering Students' Learning Strategies through Models and Modeling". I am interested in the assessment and the effectiveness of the model eliciting activities when implemented in the classroom. We are focusing on the assessment and the improvement in conceptual learning as well as problem solving. We are using a series of assessment instruments to better understand and measure the educational benefits. We are using pre- and post- concept inventories to assess improvement, an online reflection tool to assess process, and a grading rubric to assess the solution (general model and specific solution). We are identifying numerous problem solving processes used by the student teams, as well as the range of problems that

can be addressed, to determine how effective the various processes are relative to improved conceptual understanding.

Collaborators Mary Besterfield-Sacre, University of Pittsburgh, Larry Shuman, University of Pittsburgh, Karen Bursic, University of Pittsburgh, Sam Spiegel, University of Pittsburgh, Kacey G. Marra, University of Pittsburgh.

Dr. Nur Ozge Ozaltin

Nur Özge Özaltin received her master's degree and Ph.d. in the Industrial Engineering Department at the University of Pittsburgh. She received her B.S. in industrial engineering at Bogazici University in Turkey. She studies improving innovation through modeling the design process. Her methods include Bayesian network modeling, statistical and qualitative data analysis.

Dr. Mary E. Besterfield-Sacre, University of Pittsburgh

Dr. Mary Besterfield-Sacre is an Associate Professor and Fulton C. Noss Faculty Fellow in Industrial Engineering at the University of Pittsburgh. She is the Director for the Engineering Education Research Center (EERC) in the Swanson School of Engineering, and serves as a Center Associate for the Learning Research and Development Center. Her principal research is in engineering education assessment, which has been funded by the NSF, Department of Education, Sloan Foundation, Engineering Information Foundation, and NCIIA. Dr. Sacre's current research focuses on three distinct but highly correlated areas – innovative design and entrepreneurship, engineering modeling, and global competency in engineering. She has served as an associate editor for the JEE; and is currently associate editor for the AEE Journal, as well as the Advisory Board for the NAE Frontiers of Engineering Education.

Dr. Larry J. Shuman, University of Pittsburgh

Larry Shuman is senior associate dean for academic affairs and distinguished service professor of industrial engineering, Swanson School of Engineering, University of Pittsburgh. His research focuses on improving the engineering educational experience, emphasizing assessment of learning and problem solving abilities, and studying the ethical behavior of engineers and engineering managers. He was the principal investigator for a seven university NSF sponsored study on Models and Modeling that focused on using MEAs in engineering classrooms. Dr. Shuman is the founding editor of Advances in Engineering Education, and an ASEE Fellow. He holds a BSEE from the University of Cincinnati and a PhD in Operations Research from the Johns Hopkins University.

Model Eliciting Activities Motivated Problem Solving: An Analysis of Solution Paths

Abstract

As part of a larger NSF funded project focused on developing, incorporating and assessing Model Eliciting Activities (MEAs), we have investigated the solution paths student teams use in their problem solving processes. In short, MEAs require the team to develop a generalized mathematical model to solve a posed problem and to present both the solution method and the results to the "client" in a form of a written report. Our previous research found improved conceptual understanding, problem solving and professional skills obtainment when MEAs were introduced into the classroom; however, we also found that a portion of the student teams performed poorly on the activity. The particular experiment discussed here was directed at obtaining additional insight into the underlying reasons for why student teams' select incorrect solution paths when solving MEAs. In doing so, we focused on determining where and why in the problem solving process the teams started to go awry. A content analysis of the process was first used to identify underlying themes of how groups solved the problem; and then a detailed path analysis of the solution process was performed using the written reports as a dataset. In addition, data were recorded from multiple teams solving the problem 'out-loud.' The transcripts were then transcribed and analyzed using both qualitative and quantitative methods that focused on six key questions: Where do students go wrong and why? Do they iterate among solution steps (i.e., do they move from one category to another)? In iterating, do they still hold on to poor assumptions? How frequently do they correct their mistakes? Do they introduce new ideas when iterating? Finally, when new ideas are introduced, are the ideas usually correct or incorrect? Through the qualitative analysis the research team was able to gain a more in-depth understanding of why students selected certain solution paths.

1. Introduction and Background

The overall purpose of this research is to determine if the use of model eliciting activities (MEAs) in the classroom helps to improve students' ability to solve engineering problems. Model Eliciting Activities are open-ended realistic problems constructed around a few main concepts[1-8]. Originally developed by mathematics educators, and used at the pre-college level, we have been refocusing MEAs for use in upper division engineering courses. MEAs require the team to develop a generalizable, mathematical model to solve the problem and to present both the solution method and the results in a written report in the form of an engineering memorandum to a fictitious client.

As part of a larger study on models and modeling, one of our objectives was to assess the effectiveness of MEAs across various dimensions including improving conceptual learning and problem solving abilities[2,4,5]. We have implemented and assessed MEAs in the classroom to study students' problem solving, modeling and teamwork processes. When assessing the effectiveness of MEAs in improving conceptual learning and problem solving we have used three assessment methods: pre and post concept inventories (CIs) to assess learning gain, an online reflection tool to assess the problem solving process, and a rubric to assess the resulting general model and specific solution[2,4,5]. While we have explored the use of MEAs within various

engineering domains, this part of our research specifically focused on the implementation of MEAs in a two-course engineering statistics sequence. Modeling abilities play an important role in engineering. Students who are learning to model often use a variety of representation to express, test, revise and communicate their own thinking[9]. In an earlier study, we first observed the impact of MEAs on conceptual understanding. Our results strongly suggested that when used correctly, MEAs can be effective in reinforcing and integrating course concepts as well as increasing student knowledge and understanding of various professional and procedural skills (i.e., certain of the ABET eleven outcomes)[1-6]. We observed that introducing the MEAs into the classroom resulted in substantially better student performance on the post-concept inventory tests; in addition, we found students to be more confident in their answers especially for the MEA targeted concepts[3].

Over three academic years (20009-10, 2010-11 and 2011-12), three statistics MEAs were used in engineering course Probability and Statistics for Engineers 1 (ENGR 20). This course is offered each semester with certain sections designated for "majors" only (industrial engineers); the others are open to all engineering students. This enabled us to have both experimental (where the MEAs were introduced) and comparison (with no MEAs) sections. The CIs were given at the beginning and end of the semester in both the experimental and comparison sections. Hence, this concept inventory, administered pre and post, measured conceptual improvement over the duration of the course. It consisted of a subset of 20 items from two pre-established Statistics Concept Inventories. The selected questions addressed the MEA targeted concepts. The pre and post CI tests were identical. A brief overview of each MEA and the targeted technical concepts are available at www.modelsandmodeling.net along with a collection of relevant papers and MEAs developed by the seven university partners in the larger study. The analysis, methodology and results are presented in other papers[1-2].

The student teams' MEA reports provide another way to assess the MEA implementation. Each team report documents the level of understanding of the targeted concepts, and whether the concepts were used correctly. In our previous work, we found that the majority of student teams did not obtain the correct solution when solving the MEAs, even though the concepts were first presented in class, prior to the MEA being assigned, and the MEA was specifically used to reinforce those concepts [could reference earlier paper by Pinar]. The team reports, written in memorandum format, clearly documented that the various teams reached different conclusions; i.e., obtained different solutions. While each report revealed the team's proposed solution, and the path (i.e., solution process) chosen to reach that conclusion, it was difficult to ascertain why an inappropriate path was selected, nor how the team determined that one path was better than another, especially in situations when students were either not using or using inappropriately the MEA's targeted concepts. Consequently, we identified two issues requiring further attention: 1) Understanding the problem solving process students used while working on MEAs and 2) testing and documenting the actual learning benefits of MEAs.

2. Theoretical Framework

Thus, the objective of this study was to better understand where and why in the problem solving process certain teams selected an incorrect solution path. To do this, we used grounded theory coupled with a prescribed path analysis of the process to tease out the underlying problem solving themes. We did this by more closely examining how student teams addressed two statistics MEAs (Tire Reliability and Test Leads). In order to obtain the necessary data, we had

the teams solve the MEAs "think aloud." We then collected and analyzed these discussions to better determine the steps the team took to reach their final solution.

It is this additional, "think aloud" data that has permitted further understanding of the problem solving process. We wanted to determine where and why in the problem solving process the teams started down an incorrect solution path. By performing the content analysis of the process, we were able to identify underlying themes of how groups were solving the posed problems. We selected two types of problem solving processes – one more macro in nature and one that provided a finer grain analysis of the problem solving process (or micro).

For the Macro problem solving process, we adapted categories used by Atman et al and Adams and Atman as the framework for our coding[10,11]. Specifically, for the macro problem solving process, the major processes were: Problem Identification, Collecting Information, Analyzing Data, Model/Solution Formulation, Evaluation/Revise Solution and Documentation. For the micro problem solving process, the processes were: Express, Test, Revise (Change), Revise (Do Not Do Anything) and Revise (Update). This later problem solving process was developed by Kelly et al and Kelly was further augmented by Hamilton[12,13]. Thus we performed a path analysis of the solution process; we collected data through a series of designed experiments and analyzed the common patterns and data using qualitative methods. Figure 1 provides an overview of the theoretical framework employed.
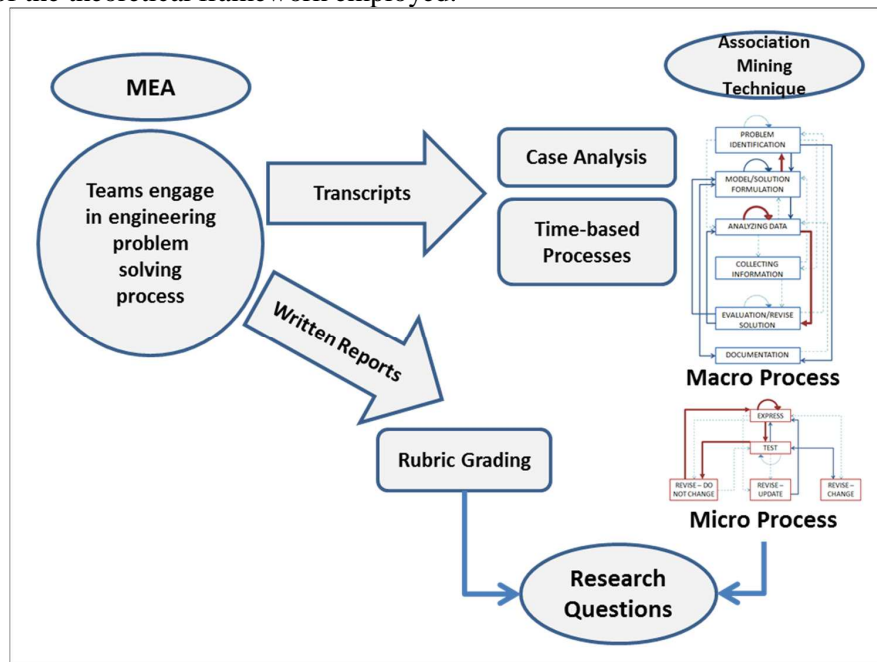

Figure 1. Theoretical Framework for Study

In this paper we utilized the framework to address two primary research questions:

1. Where do students go wrong and why? Given this, as part of their solution process, do student teams adopt proper assumptions or do they adopt and hold on to poor assumptions?

2. Do they iterate towards an engineering solution? Intuitively we know students do iterate, but specifically how many times do they iterate; and do they correct their mistakes, introduce new ideas, and are these new ideas appropriate?

From these questions, we gained additional insight in determining solution patterns across the teams, as well as differences between the two MEAs employed, as one was more difficult to resolve (Tire Reliability) than the other (Test Leads).

3. The Experiment

Sophomore Industrial engineering students take a two-course sequence in Probability and Statistics for Engineers. During the 2012 Spring Term 72 industrial engineering students were enrolled in the second course. We solicited volunteers (as part of a team) from that course to solve one of the two MEAs.

As noted, the two MEAs used were Tire Reliability and Test Leads (see www.modelsandmodeling.net for the MEAs). Each MEA consisted of two parts: an individual part and a group part. The individual part comprised of several preparatory questions that encouraged students to think about aspects of the larger problem. Hence the individual assignment served as a preliminary guide to the main assignment. All subjects had completed the MEA CNC Machine as part of a class assignment, so they were familiar with the MEA protocol, requirements, and expectations. In addition, all students completed both the pre concept inventory at the beginning of the course and the post concept inventory at the end of the semester. As noted, the statistics inventory used in these studies consisted of 20 multiple choice items selected from two established and tested concept inventories that addressed the concepts around which the MEAs had been constructed [14-16].

In total, 27 student volunteers were recruited. These were randomly divided into nine teams (three students each), and each team was randomly asked to solve one of the two MEAs. Each MEA took approximately three hours to resolve; all conversations were recorded during this solution time. After completing the individual part, the team met in a specially equipped room enabling us to record their conversations while solving the MEA. They were also asked to prepare a written report describing their recommended solution, as well as complete an individual reflection questionnaire related to the team's solution process. They were provided with a computer to write up their results in memorandum format for the simulated client. In addition to the recording equipment and computer, the room also contained a statistics textbook and calculators.   Students were instructed to think aloud and talk through the process. Upon completion of the MEA and the written report, the students individually completed the reflection exercise. All but one of the nine teams submitted a written report; and as a result, we report on only eight teams in this paper. The students were paid for their time; the experiment had human subjects' approval.

The team's verbal responses were transcribed by the University Center for Social and Urban Research (UCSUR) and subsequently analyzed. Reflection data were coded by individual and team. Prior to performing the experiment, two pilot teams solved one of the MEAs. These pilot transcripts were later used to test the coding scheme developed for the qualitative analysis. Using the theoretical framework provided, content analysis of the process was conducted on each transcript to determine the underlying themes of how the teams were solving the problem along

the two problem solving models (micro and macro). Case studies were also developed for each of the teams based on the transcripts. The content analysis and case study analysis was performed by different researchers so that findings could be compared in an unbiased manner.

Written MEA reports were graded according to the scoring rubric created for each MEA. Quantitative analysis consisted of the analysis of the written reports (grading) and comparing the pre and post concept inventory scores (to determine if there is the difference in the gain obtained by students exposed to MEAs with the rest of the class). Specific questions related to MEA targeted concepts were also analyzed. These results are reported elsewhere.

4. Data Analysis and Methodology

4.1 Qualitative data analysis methodology

We performed a qualitative analysis of the transcribed data[17] to obtain an in-depth understanding of students' problem solving process for the nine experimental teams. As mentioned, preliminary exploratory analysis was performed on pilot transcripts to make connections to the research questions; and a coding scheme with categories was developed for the experimental transcripts. Each team's transcript was reviewed and coded under both macro and micro processes. For each team, a macro and micro process report was prepared.

After building the time-based macro and micro processes, an association mining technique was used to discover the degree of associations among the categories in these processes. We determined the associations by counting the number of times category X is followed by category Y for each team during the project timeline. The details of the coding are summarized in the coding handbook (please contact authors for a copy of the handbook). Testing of the coding scheme was performed on two pilot transcripts, which were not included in the final analysis. One researcher, who had expertise in qualitative coding, coded all the transcripts; and the coding consistency was determined via a coding handbook. Although only one researcher coded the transcripts, the coding handbook was reviewed by two researchers.

Process maps were developed from the coding to better draw conclusions. First, association mining analysis was applied to see how often a team moved from one category to another; i.e., from category A to category B. After the frequencies of these iterations were determined, their strengths were determined based on natural break points (i.e, the frequencies were sorted enabling the natural break points to be identified). Three levels of strength were specified: weak, moderate and strong. This enabled maps to be drawn based on the strength of the associations. Here, iteration is defined as the movement from one category to another. This movement depicts the directional relationship between design categories. Furthermore, the frequency of these movements over time period provides information on the strength of the relationships.

Since the research objective was to understand how students solved these problems, their behavior (i.e, the iterations among the categories) was determined, which then yielded the solution paths. Hence, the association mining technique enabled us to identify the iterations in macro and micro processes[18]. By doing this, we could determine the students' paths used when solving the problem. We could also analyze the relation between the success of the solution process and the path followed to determine if a more general solution path exists.

The association mining technique identifies the relationships between variables; therefore relationships in the form of association rules or sets of frequent items can be uncovered. An association rule is an implication expression of the form X → Y, where X and Y are disjoint events, i.e., X ∩ Y = Ø. The strength of an association rule is measured by its support and confidence[18]. Support is the probability that two different categories appear consecutively throughout the design process; and confidence is the conditional probability that a particular category occurs after a given category. These probabilities are used to determine the degree of randomization in the data. Maps of the various associations are then created based on support probability to determine how teams iterate among the various macro and micro processes.

After building the time-based macro and micro processes, the association mining technique was used to discover the degree of associations among the categories in these processes. We examined the associations by counting the number of times category X is followed by category Y for each team during the project timeline. Table 1 presents an example calculation of the support probabilities. This example is from Team C (Test Leads MEA) in which the macro process had 56 movements. For the team considered in the example, "analyzing data" appears twice after "problem identification" during the timeline; so the support probability from "problem identification to analyzing data" is 2/56 =0.04.

Table 1: Calculation of Support Probabilities – Team C – "Test Leads"

From Category X	To Category Y	Number of Movements	Support Probability
Problem identification	Problem identification	6	0.11
	Analyzing data	2	0.04
	Model/solution formulation	3	0.05
	Evaluation/revise solution	2	0.04
	Collecting information	1	0.02
	Documentation	1	0.02

All calculated probabilities for Test Leads and Tire Reliability teams were separately sorted. Based on the natural break points, the degrees of associations (i.e. strong, moderate, and weak) were determined. These are given in Figure 2 for macro processes and Figure 3 for micro processes. Note that for the macro processes, natural breakpoints were evident, while for the micro processes, there were no natural breakpoints. Consequently, for the micro process a subjective classification was applied (i.e., if $p \leq 0.03$, then the association is weak; if $0.03 \leq p \leq 0.09$, then the association is moderate; and if $p \geq 0.09$, then the association is strong).
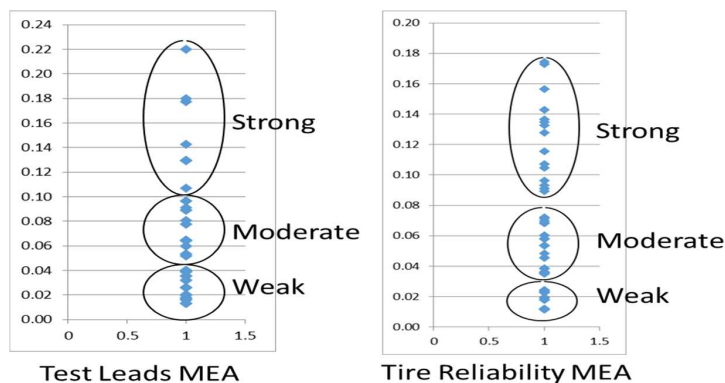

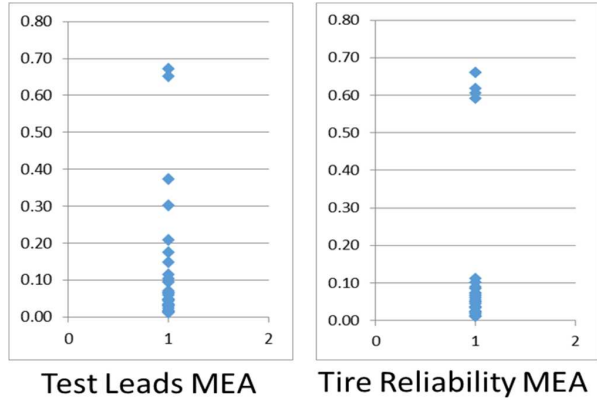**Figure 2: Sorted Probabilities for Macro Processes**

Figure 3: Sorted Probabilities for Micro Processes

These associations were then graphed onto maps, such as shown in Figure 4. If the association is strong, then a bold and thick (red) arrow is used to demonstrate the particular association; if it is moderate, then a bold arrow is used; and if the association is weak, then a dashed arrow is used. For example, in the figure, there is a strong association from "problem identification" to "problem identification", and a moderate association occurs from "model/solution formulation" to "analyzing data". A weak association is observed from "analyzing data" to "collecting information".
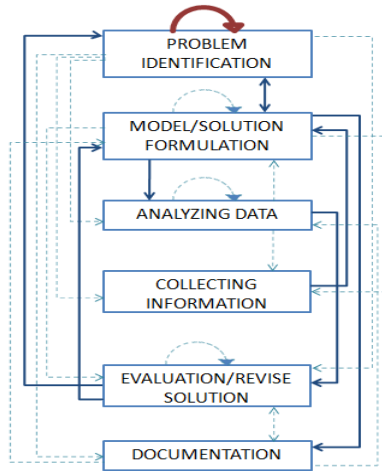
Figure 4: Example Association Map of the Macro Solution Process (showing strengths of the associations)

4.2 Quantitative Analysis

We evaluated the data collected through written reports of the two MEAs and concept inventory scores. Written reports were graded according to the developed grading rubrics (rubrics available on www.modelsandmodeling.net). We also compared pre and post concept inventories for the whole class and for the students exposed to MEAs vs. students who did not solve MEAs. The grades of the written reports are provided in this paper; and the concept inventory results will be provided in a forthcoming paper.

5. Results and Discussion

After preparing maps for all teams, we developed comparison matrices for both Micro and Macro processes for both groups of teams: Test Leads MEA and Tire Reliability MEA, as depicted in Tables 2 through 5. For each of the two MEAs, we discuss how each team migrates along the two sets of problem solving processes and how these migrations potentially relate to their performance on the MEA as measured by their scores on the written report rubric. As mentioned, teams were randomly assigned to one of the two MEAs. Four teams (B, C, F and G) were assigned to the Test Leads MEA; and five teams (A, D, E, H, and, J) were randomly assigned to the Tire Reliability MEA.

5.1 Test Leads MEA

Tables 2 and 3 provide the Macro and Micro processes for the teams that conducted the Test Leads MEA, respectively. Each table provides the strength of the associations between the different categories in the two processes we employed (macro and micro). For each team, we provide an account for how the team progressed through the problem and relate the strength of their associations to how they potentially solved (or did not solve) the MEA task. Each team is discussed in order of their rubric score (highest score to lowest score). Team F had the highest rubric score on the MEA; and Team B had the lowest score on the MEA. Tables 2 and 3 also include the total number of iterations per team.

Team F (Rubric score: 96) initially had a misunderstanding about the Central Limit Theorem (CLT), but during the process they corrected themselves. The team's calculations were correct (mean, and the standard deviation). Further, they verified whether the data was uniformly distributed or not, even though that information was given in the problem statement. The team was confused about the width of the confidence interval provided; and even though the textbook was provided and they could find the correct formula to calculate the sample size, they still tried to justify that their result was correct. This team also considered sample size, and the "rule-of-thumb" sample size of $n=30$ (from CLT), as well as much smaller sample sizes of n=10 and 12 (their quotation related to 10 and 12: "*In order to validate this sample size of 21 leads, we investigated the normality of distribution of means for a sample size of 21 (and smaller sample sizes of 10 and 12)*". As a result they decided to use 21 and continued the tests with $n=21$.

Thirteen percent of iterations were between "analyzing data" to "evaluation/revise solution"; and this was the only strong association in the macro process map for Team F as shown in Table 2. While this strong association was one-sided for this team, it was two-sided for both of the lowest graded teams (Teams G and B – to be discussed). Team F exhibited a weak association from "problem definition" to "problem definition." This association was also weak for the lowest graded team (Team B); in contrast, for the two middle teams, it was moderate or strong (Teams C and G). Team F also had a strong association related to documentation. For the micro process (Table 3), this team did not have any associations related to "revise – change". The strongest association from "test" to "revise" is specifically to "revise – update" (for all other teams, that association is from "test" to "revise- do not change").

Table 2: Comparison Matrix (Macro Processes): Test Leads MEA

From Category X	From Category X *Total Iterations*	Team F *46*	Team C *46*	Team G *34*	Team B *44*
Problem identification	Problem identification	weak	**strong**	moderate	weak
	Model/solution formulation	moderate	moderate	moderate	moderate
	Analyzing data	weak	weak	weak	weak
	Collecting information	weak	weak		
	Evaluation/revise solution	weak	weak	weak	
	Documentation	moderate	weak		
Model/solution formulation	Problem identification	moderate	moderate	weak	moderate
	Model/solution formulation	moderate	weak	weak	**strong**
	Analyzing data	moderate	moderate	weak	moderate
	Collecting information		weak		
	Evaluation/revise solution		weak	moderate	moderate
	Documentation	weak	moderate		weak
Analyzing data	Problem identification			weak	weak
	Model/solution formulation	weak	weak	weak	moderate
	Analyzing data	moderate	weak	weak	
	Collecting information		weak		
	Evaluation/revise solution	**strong**	moderate	**strong**	**strong**
	Documentation	weak		weak	
Collecting information	Problem identification				
	Model/solution formulation		moderate		
	Analyzing data	weak			
	Collecting information				
	Evaluation/revise solution			weak	
	Documentation	weak			
Evaluation/revise solution	Problem identification		moderate	weak	weak
	Model/solution formulation	weak	moderate	weak	moderate
	Analyzing data	moderate		**strong**	**strong**
	Collecting information	weak			
	Evaluation/revise solution		weak	weak	moderate
	Documentation	weak	weak	weak	weak
Documentation	Problem identification	moderate			weak
	Model/solution formulation	moderate	weak		weak
	Analyzing data	weak	weak		
	Collecting information				
	Evaluation/revise solution	weak	weak	weak	
	Documentation				

Team C (Rubric Score: 77) expressed a lot of confusion about the parameters of the assigned problem. They were confused about the width of the confidence interval and the standard deviation. They believed that they were supposed to assume a normal distribution, so they discussed how a uniform distribution behaves like a normal distribution. They concluded that "*It's uniform throughout, so like, that translates to the bell curve*". They also stated that *"Since the measure of interest, diameter, appears to be uniformly distributed, we were able to assume that the measurements followed a normal distribution."* Instead of calculating the standard deviation, they used 0.12, and then they used a different value. They applied a z-test, but they did not refer to the CLT. In general, this team was confused and did not know how to proceed. Overall, they did not understand which test to apply (when and why those tests are used). As shown in Table 2, eleven percent of the iterations were from "Problem identification" to "Problem identification", and this was the only strong association for their macro process. This

further supports the fact that they were confused and did not know how to approach the problem (i.e., they were constantly going back to problem definition). For their micro process (see Table 3), this team had the lowest number of iterations.

Table 3: Comparison Matrix (Micro Processes): Test Leads MEA

From Category X	To Category Y	Team F	Team C	Team G	Team B
	Total Iterations	*22*	*21*	*39*	*29*
Express	Express	**strong**	**strong**	**strong**	**strong**
	Revise-do not change	weak	weak	weak	weak
	Revise-update			weak	weak
	Revise-change			weak	weak
	Test	**strong**	**strong**	**strong**	**strong**
Revise-do not change	Express	weak	moderate	**strong**	moderate
	Test			weak	weak
Revise-update	Express	weak		moderate	weak
	Test	weak			
Revise-change	Express		weak	weak	weak
	Test			moderate	
Test	Express	moderate	moderate	moderate	**strong**
	Revise-do not change	weak	moderate	**strong**	moderate
	Revise-update	moderate		weak	weak
	Revise-change			moderate	weak
	Test	weak		weak	

Team G (Rubric Score: 65), like Team C had some difficulty understanding the problem. In the beginning, they believed that the data becomes normal, but then one of the members corrected the team. They considered the CLT; however, their implementation was not appropriate. They were confused about the differences in confidence levels (i.e., 95% vs. 99%). The team did apply the correct formula for the confidence interval (CI), however the values they used were not correct. They applied the correct formula for standard deviation, but again their values were incorrect. In the end, the team never fully determined what they needed to do to resolve the MEA.

For this team two strong associations were documented. As shown in Table 2 the first was from "analyzing data" to "evaluation/revise solution" (17%); and the second strong association was from "evaluation/revise solution" to "analyzing data" (21%). These two overriding movements indicate that the team was confused; this type of relationship is documented for the two teams that had the lowest grades on this particular MEA. Further, this team had a moderate association from "problem definition" to "problem definition;" further indicating that they had trouble understanding the initial problem as they were often going back to determining what the problem actually was. The team even indicated: *"Well, we really had no idea what we're going to do at this point, so we took a shot in the dark."* From this point they decided to start over; and went back to the problem statement and corrected their mistakes. Finally, there were no strong or moderate associations related to documentation; further indicating that the team did not document their solution properly. This is as well true for the lowest graded team. For the micro process, team G went to "revise" significantly more than the other teams, as shown in Table 3

(i.e., "Revise – change, Revise – do not change, Revise – update"). There is also a strong loop for "Express-test-revise-do not change-express") which indicates that even though they decided not to change, they went back and re-tested as they were not satisfied with the answer obtained. This team is the only team with the strongest (moderate) two-sided association between "test" and "revise-change". The strongest association from "test" to "revise" is specifically to "revise – do not change."

Team B (Rubric Score: 45) had some difficulty understanding the problem. They were confused about the CI (95% vs. 99%). The Team assumed that data is normally distributed instead of the correct uniform distribution. They mentioned the correct formula for the sample size at first, but ended up using an incorrect formula. Further, they did not discuss the CLT. As shown in Table 2, the largest proportion of their iterations was from "Analyzing Data" to "Evaluation/revise solution" (19%), with 11% from "Evaluation/revise solution" to "Analyzing data." Also, 11% were from "Model/solution formulation" to "Model/solution formulation". Interestingly, this team did not use "Collecting information". They did not have any strong or medium associations related to "Documentation," which is likely linked to the poor written rubric grade. For their micro process (Table 3), this team has the second highest number of iterations. Team G, had the highest number of iterations. The strongest association from "test" to "revise" was specifically to "revise – do not change." This was consistent for all teams except for the highest graded team F.

5.2 Tire Reliability MEA

Similarly to Test Leads, Tables 4 and 5 provide the Macro and Micro processes for the teams that conducted the Tire Reliability MEA, respectively. Each table provides the strength of the associations between the different categories in the two processes we employed (macro and micro); and for each team we provide an account of their progressions along the processes relating the strength of associations to their rubric score. Each team is discussed in order of their rubric score (highest score to lowest score). Team E had the highest rubric score on the MEA; and Team A had the lowest score on the MEA. Note that Tire Reliability is considerably harder, less straightforward, MEA resulting in lower rubric grades compared to the Test Leads MEA. Table 5 also includes the total number of iterations per team.

Team E (Rubric score: 71) assumed that the golden standard data was normally distributed without conducting appropriate tests to make such a claim. Further, they did not know what reliability means, or how to calculate it. To find the acceptable reliability range, the team applied a box-and-whisker plot, and found that the range was 0.8-1.26. During the problem solving process, they logically evaluated the data (e.g., they found that there might be some "failed tires" even if the set is reliable, etc.); and they discovered that the given data set of 100K had no failures, but unfortunately then misinterpreted the information.

Overall, this team was confused about the problem. Consequently, they applied a number of hypothesis tests; and conducted trial-and-error testing. The team's largest number of iterations (16%) was from "Analyzing data" to "Evaluation/revise solution"; another 12% were from "Model/solution formulation" to "Analyzing data," as shown in Table 4. For both processes, this team had the highest number of iterations. Table 5 shows that the strongest association was from "test" to "revise" (specifically to "revise – do not change"). Finally, their report was detailed, well-written, but implemented an incorrect procedure.

Team D (Rubric Score: 48) also did not know how to calculate reliability or fundamentally what it meant. They assumed that the gold standard data was distributed normally without plotting or developing a histogram. Further, the team expressed that they were really confused about a number of points (e.g., alpha levels, tests, procedure). Finally, while interpreting their test results they voiced concerns related to their confidence of their decision. Given their confusion, this team turned to trial-and-error to solve the MEA. As shown in table 4, 17% of their iterations were from "Model/solution formulation" to "Model/solution formulation", 13% from "Model/solution formulation" to "Analyzing data," 12% from "Evaluation/revise solution" to "Model/solution formulation," and 10% from "Analyzing data" to "Evaluation/revise solution." This team demonstrated four strong associations in their macro process. In addition, there was a strong loop from "model/solution formulation" → "analyzing data"→ evaluation/revise solution → model/solution formulation. This is potential indication that the team was not confident in their procedure. For their micro process (Table 5), this team demonstrated the second highest number of iterations.

Team H (Rubric Score: 48) also believed that the distributions of the data sets should be normal; otherwise none of the statistical tests could be applied. They assumed normality because they had a large number of data points. In the transcribed data, the students expressed a lot of confusion and low confidence about what to do and how to approach the problem. The team decided to use the z-test, ANOVA and F-test to address the requirements of the problem. Overall, this team did not know how to calculate reliability, and how to compare the samples to the "gold" one. Table 4 indicates the largest portion of iterations (14%) were from "Analyzing data" to "Evaluation/revise solution"; 9% were from "Model/solution formulation" to "Problem identification" and 9% were from "Analyzing data" to "Analyzing data". There was a strong association from "model/solution formulation" to "problem identification." The direction of the arrow was expected to be opposite, but there was a moderate association from "problem identification" to "model/solution formulation." For their micro process (Table 5), the team demonstrated many weak associations compared to the other teams, but did shown a strong association from "express to express" as did other teams. This team had the second lowest number of iterations.

Table 4: Comparison Matrix (Macro Processes): Tire Reliability MEA

From Category X	From Category X	Team E	Team D	Team H	Team A
	Total Iterations	*53*	*51*	*49*	*52*
Problem identification	Problem identification	weak	moderate	weak	weak
	Model/solution formulation	weak	weak	moderate	moderate
	Analyzing data	moderate		weak	
	Collecting information	weak		weak	
	Evaluation/revise solution	moderate	moderate		moderate
	Documentation	weak		moderate	moderate
Model/solution formulation	Problem identification	weak	weak	**strong**	moderate
	Model/solution formulation	weak	**strong**	moderate	**strong**
	Analyzing data	**strong**	**strong**	moderate	moderate
	Collecting information		weak		
	Evaluation/revise solution	weak	moderate		weak
	Documentation	weak		moderate	weak
Analyzing data	Problem identification				
	Model/solution formulation	moderate	moderate	weak	weak
	Analyzing data	moderate	weak	**strong**	weak
	Collecting information		weak	weak	

From Category X	From Category X	Team E	Team D	Team H	Team A
	Total Iterations	*53*	*51*	*49*	*52*
	Evaluation/revise solution	**strong**	**strong**	**strong**	**strong**
	Documentation	weak			
Collecting information	Problem identification		weak		
	Model/solution formulation	weak	moderate	weak	
	Analyzing data				
	Collecting information				
	Evaluation/revise solution	weak	weak	weak	
	Documentation				
Evaluation/revise solution	Problem identification	moderate	weak	weak	**strong**
	Model/solution formulation	moderate	**strong**	moderate	moderate
	Analyzing data	moderate	weak	moderate	**strong**
	Collecting information	weak	moderate		
	Evaluation/revise solution	weak	moderate	weak	weak
	Documentation	moderate	weak		moderate
Documentation	Problem identification	weak			weak
	Model/solution formulation	weak		moderate	weak
	Analyzing data	weak		weak	
	Collecting information				
	Evaluation/revise solution	moderate			moderate
	Documentation				

Team A (Rubric Score: 39) divided the work across the team; one member was calculating the mean and standard deviation while the others read the assignment and performed other tasks. This team misinterpreted the histograms they created as they concluded that all the data sets were normally distributed. They, as well, did not know how to calculate reliability; and also expressed confusion about the overall assignment.

From Table 4, the largest number of their iterations (14%) were from "analyzing data" to "evaluation/revise solution"; 9% were from "Model/solution formulation" to "Model/solution formulation"; another 9% were from "Evaluation/revise solution" to "Problem identification", and a third 9% were from "Evaluation/revise solution" to "Analyzing data." They had four strong associations in their macro process. There was a strong association from "Evaluation/revise solution" to "Problem identification". The direction of the arrow was expected to be in the opposite direction (but there was a moderate association from "Problem identification" to "Evaluation/revise solution"). Interestingly, this team did not use any "Collecting information" in their macro process. There was a loop in the micro process (consisting of moderate and strong associations) among express → test →revise-do not change →express. For their micro process (see Table 5), this team had the lowest number of iterations, which corresponds with the lowest grade on the written report. Further, this team did not use any "Revise-update" or "Revise-change" connections in their analyses.

Table 5: Comparison Matrix (Micro Processes): Tire Reliability MEA

From Category X	From Category X	Team E	Team D	Team H	Team A
	Total Iterations	*26*	*25*	*23*	*19*
Express	Express	**strong**	**strong**	**strong**	**strong**
	Revise-do not change	**strong**	weak		weak
	Revise-update	weak		weak	
	Revise-change	moderate	moderate		

From Category X	From Category X *Total Iterations*	Team E *26*	Team D *25*	Team H *23*	Team A *19*
	Test	weak	moderate	**strong**	moderate
Revise-do not change	Express	**strong**	moderate	weak	moderate
	Test		weak		weak
Revise-update	Express	weak		weak	
	Test				
Revise-change	Express	moderate	moderate	moderate	
	Test				
Test	Express	weak	moderate	moderate	moderate
	Revise-do not change	weak	moderate	weak	moderate
	Revise-update				
	Revise-change			moderate	
	Test				

6. Conclusions and Future Research

In this paper we utilized the framework to address two primary research questions:

> 1. Where do students go wrong and why? Given this, as part of their solution process, do student teams adopt proper assumptions or do they adopt and hold on to poor assumptions?

> 2. Do they iterate towards an engineering solution? Intuitively we know students do iterate, but specifically how many times do they iterate; and do they correct their mistakes, introduce new ideas, and are these new ideas appropriate?

Test Leads MEA

For the Test Leads MEA, the highest graded Team has the lowest association from "Problem identification" to "Problem identification" which may indicate high level of understanding as well as adopting proper assumptions in the early stages of the solution process. This as well shows that the Team is confident in their approach. Lower graded teams have stronger associations in the same category showing initial confusion and not understanding the problem; which may indicate lack of knowledge and poor assumptions. All teams have a "moderate" association from "Problem identification" to "Model Formulation" indicating the existence of iterations and review at the beginning of the solution process.

All teams have "moderate" association form "Model Formulation" to "Evaluation/revise solution" which indicates that they were revising and reviewing in order to properly formulate the problem.

All teams have "strong" association form "Analyzing data" to "Evaluation/revise solution" which demonstrates that they went through several phases when analyzing data in order to properly understand the given information. From the micro level analysis we can get the deeper understanding if the "revise solution" part resulted in the revise-do not change or revise –update.

For the two lowest graded teams, there is a "strong" association from "Evaluation/revise solution" to "Analyzing data" indicating possible dissatisfaction with the obtained solution. Due to possible lack of knowledge these teams were unable to correct themselves.

On the "micro" level, there is a "strong" association for all teams from "Express" to "Express", indicating the high number of iterations. There is a "strong" association from "Express" to "Test" and "moderate" association from "Test" to "Express". Based on this information, we can conclude that students iterate, express ideas and test ideas; however we cannot conclude that they are correcting their mistakes or assumptions. Only the highest graded Team has a "moderate" association form "Test" to "Revise-update". Two lowest graded teams have a weak association, meaning that they are not updating their solution, again possibly due to the lack of knowledge. In addition, only the highest graded Team has a weak association from Test to Revise-do-not-change; other three teams have moderate and strong associations, indicating that most students revise but do not correct their mistakes.

Tire Reliability MEA

For the Tire Reliability MEA, three teams have moderate or strong association (the highest graded Team has the weak association) from "Model/solution formulation" to "Model/solution formulation" indicating the initial confusion and not understanding the problem at the early stages.

For this MEA, students have stronger associations later in the solution process, as there is a "strong" (two teams) and "moderate" association (two teams) from "Model/solution formulation" to "Analyzing data" and "strong" association from "Analyzing data" to "Evaluation/revise solution" for all teams. All teams as well have either "moderate" or "strong" association from "Evaluation/revise solution" to "Model/solution formulation". This indicates that students were going back and forth from "Model/solution formulation" to "Analyzing data" to "Evaluation/revise solution", but unfortunately did not correct their mistake in order to properly solve the problem.

On a micro level, there is a "strong" association for all teams form "Express" to "Express", confirming the initial confusion and possible misunderstanding of the problem. Two teams have "weak" and two have "moderate" association from "Test" to "Express". Similarly, there are two "weak" and two "moderate" associations from "Test" to "Revise-do not change", indicating high number of iterations, but no improvement in the solution process. Students were not revising their mistakes and poor assumptions. We did not record any association from "Test" to "Revise-update" and only one team had "moderate" association from "Test" to Revise-change".

For the Test leads MEA, which is a straightforward assignment compared to the Tire Reliability MEA, students showed initial confusion and higher number of iterations early in the process. The higher graded teams, had stronger associations early and corrected their mistakes, which resulted in the correct solution and correct solution path. Lower graded teams, had higher number of iterations, strong associations related to Revise – do not –change, which indicates dissatisfaction with their solution, but as well lack of knowledge, as they were not able to correct themselves.

For the Tire Reliability MEA, the assignment is harder and less straightforward, the teams showed higher number of iterations, indicating more confusion especially in the early phases. Stronger associations to and from "Revise-do not change" indicate dissatisfactions with the obtained solution, but being unable to adopt proper assumptions.

Acknowledgment

References

1. Vidic N., Shuman L. J., Besterfield-Sacre M., Bursic K. M., T.P. Yildirim, and N. Siewiorek (2011). "Learning Impacts Gained from Introducing Model Eliciting Activities (MEAs) in an Introductory Statistics Course", *Proceeding of the Industrial Engineering Research Conference, Reno, Nevada, May 21-25, 2011.*
2. Bursic K., Vidic N., Yildrim T. P., Besterfield-Sacre M., Shuman L., Renee Clark (2013), "The Impact of Simulating Real Life Experiences in Engineering Education: A Case for Model Eliciting Activities", to be submitted to Journal of Engineering Education.
3. Siewiorek N., Shuman L. J., Besterfield-Sacre M., Vidic N., Bursic K. M., (2012). "Students' Confidence Levels in Technical Concept Knowledge with Model Eliciting Activities", *2012 Proceeding of the American Society for Engineering Education Annual Conference,* San Antonio, TX.
4. Shuman L. J., Besterfield-Sacre M., Bursic K. M., Vidic N., T.P. Yildirim, and N. Siewiorek (2012). "CCLI: Model Eliciting Activities", *2012 Proceeding of the American Society for Engineering Education Annual Conference,* San Antonio, TX.
5. Shuman L., Besterfield-Sacre M., Bursic K., Vidic N., Yildirim P. (2011), "CCLI: Model Eliciting Activities: Experiments and Mixed Methods to Assess Student Learning – Part II")," *2011 Proceeding of the American Society for Engineering Education Annual Conference,* Vancouver, B.C.
6. Shuman, L.J., Besterfield-Sacre, M., R. Clark, Yildirim, T.P. and K. Bursic (2009) "Introducing An Ethical Component to Model Eliciting Activities", *2009 American Society for Engineering Education National Conference*, Austin, TX, June 14-17, 2009.
7. Shuman, L., T. Moore, M. Besterfield-Sacre, H. Diefes-Dux, E. Hamilton, R. Miller, B. Olds, and B. Self, "Improving Engineering Students' Learning Strategies Through Models and Modeling," 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY, October 22-25, 2008.
8. Bursic, K., Shuman, L.J., and Besterfield-Sacre, M., "Improving Student Attainment of ABET Outcomes Using Model Eliciting Activities (MEAs)", 2011 Proceedings of the American Society for Engineering Education Annual conference and Exposition, Vancouver, British Columbia.
9. Moore T. J, Miller R. L., Lesh R. A., Stohlmann M. S., and Kim Y. R., (2013), "Modeling in Engineering: The Role of Representational Fluency in Students' Conceptual Understanding", Journal of Engineering Education, January 2013, Vol. 102, No. 1 pp.141 – 178.
10. Atman, C. J., J. R. Chimka, K. M. Bursic and H. L. Nachtmann, 1999, "A Comparison of Freshman and Senior Engineering Design Processes" Design Studies Vol. 20 No 2 (pp 131-152)
11. Adams, R. and C. J. Atman, 1999, "Cognitive Processes in Iterative Design Behavior" Proceedings of the Annual Frontiers in Education Conference, November, San Juan, Puerto Rico.
12. Kelly, A.E., R.A. Lesh, and J.Y. Baek (Eds.). (2008) *Handbook of design research methods in education: Innovations in science, technology, engineering and mathematics learning and teaching.* New York: Routledge.
13. Lesh, R., E. Hamilton and J. Kaput (Eds.). (2007). *Foundations for the future in mathematics education.* Mahwah: Lawrence Erlbaum Associates.

14. Allen, K., Stone, A., Reed-Rhoads, T. and Murphy, T.J. "The Statistics Concepts Inventory: Developing a Valid and Reliable Instrument", ASEE Conference, 2004.

15. delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). "Assessing students' conceptual understanding after a first course in statistics", *Statistics Education Research Journal, 6*(2), 28-58.

16. Allen, K. Reed-Rhoads, T. and Terry, R. "Work in Progress: Assessing Student Confidence of Introductory Statistics Concepts", Frontiers in Education Conference, San Diego, CA, 2006.

17. Miles, B. M. and Huberman, 1994, "Qualitative Data Analysis: An Expanded Sourcebook" Sage Publication Inc., 2nd edition18. Tan, P., M. Steinbach, and V. Kumar, 2006, "Introduction to Data Mining" Addison-Wesley, Boston