2005-1601

Multi-section Freshman Classes with Laboratories: Lecture as Intro vs. Lecture as Wrap-up

Jon Sticklen, Mark Urban-Lurain, Timothy Hinds
Taner Eskil, Marilyn Amey

Michigan State University

Introduction

A common instructional model for freshman engineering is the lecture/laboratory model. In this model, students usually spend two to four hours per week in a large lecture section typically of one hundred or more students, and three to six hours per week in small laboratory (or recitation) sections typically of twenty or fewer students.

Although not universal, the most common implementation of this instructional model is that lecture introduces material of a given "unit" while laboratory (or recitation) sections are used to provide hands on, detailed experience with applying knowledge introduced in assigned readings and lecture. The paradigm on which this implementation is rooted would be along the lines that students need a framework for understanding before they can apply material of a given unit, and that such a framework is best developed by students reading assigned material then hearing a professor go over the same material to emphasize important points.

There is a critical flaw in the standard lecture-before-lab implementation: it depends on students reading assigned material before lecture. If not, then lecture is unintelligible to students who have not read the assigned readings supposing the instructor hits the "high" or "hard" points of a unit, or lecture becomes a replacement for assigned readings supposing the instructor simply "plows through" material from the assigned readings. Neither of these two results is desirable, and neither places the freshman learner in a position to actively engage in her own learning.

An alternative implementation of the "large lecture/lab" instructional model would reverse the order of lecture and lab (or recitation). Students would be expected to read material, attend laboratory sections emphasizing hands on work, then at the end of the "cycle" students would attend lecture. The lecture in this implementation would play the role of "wrap up" for the unit students have completed - including making generalizations from the specifics students have learned, and demonstrating any common mistakes students make when applying the material of the unit. In this implementation, students are more or less obligated to read assigned material before their first unit meetings (labs/recitations) because in lab they must "perform" using what they have learned from the assigned readings.

The bottom line would be that it's easy for a student to "hide" in lecture, but not in small lab sections. The pedagogically larger picture is that "lecture as wrap up" should require students to take more responsibility for their own learning, and in the end be more actively engaged in their learning than the more common "lecture as introduction" path.

In fall semester, 2004, we undertook an experiment to objectively compare the "lecture as wrap up" implementation to the more standard "lecture as introduction" implementation. Although not included in our initial focus, in the course of analysis of our data we found it necessary to consider the possible effects of multiple underlying populations on our data samples.

In this report, we describe this experiment, the results of the experiment, an analysis of our results, and the implications of our results.

Background and major hypothesis

One of the current bedrocks of pedagogy is active learning and its importance in transforming the educational enterprise from a view of the student as a vessel into which the professor pours "knowledge" to one in which the learner is actively engaged in her own construction of knowledge. [1] One example of the introduction of principles of active learning into engineering studies can be found in [2]. The goal of establishing active learning has become wide spread in computer science and engineering to the extent of enabling students to set the term grade they desire, then work towards that goal. [3].

Pursuit of active learning is one of the backdrops for research reported here. As noted above, we believe that requiring students to do assigned reading before any class dependent on the assigned reading has the effect of actively engaging students in their learning process, certainly more so than the standard lecture situation in which lecture material closely mirrors assigned reading.

Specifically, our experimental hypothesis is that students who participate in *lecture as wrap up* will perform better than those who participate in *lecture as introduction*. Any past studies on this specific issue were extremely difficult to find, and in fact, we found no relevant literature.

Description of the experiment

Computer Science and Engineering (CSE) 131 is a high enrollment (approximately 250-300 students per term), multi-section (approximately 24 sections), freshman engineering course in technical problem solving with MATLAB. It is offered fall and spring semesters with an additional offering in summer term with a substantially lower enrollment. CSE 131 is a required gateway course for most majors in the College of Engineering, Michigan State University. The standard "Calculus 1" is a pre-requisite/co-requisite for CSE 131.

Because of a scheduling "glitch," in fall semester, 2004, CSE 131 was offered in two lecture sections at opposite ends of the week. One lecture section met on Monday nights at 7:00 p.m., with associated labs running on Tuesdays, Wednesdays, Thursdays, and Fridays. A second lecture section met on Friday mornings at 10:20 a.m., with associated labs running on Tuesdays, Wednesdays, Thursdays, and Fridays. Lecture section enrollments were approximately the same, with 122 students enrolled in the Monday lecture section at the end of the semester and 109 students enrolled in the Friday lecture section at the end of the semester. Laboratory sections entailed enrollments of a maximum of 16 students, and were not mixed – that is laboratory sections included students either totally in the Monday lecture section, or totally in the Friday lecture section.

Each student in CSE 131 meets for one lecture session per week lasting one hour and twenty minutes, and meets twice per week in laboratory sessions twice per week with each lab meeting lasting one hour and twenty minutes. Thus the Monday lecture session students met in lecture *before* participating in any lab assignments, while Friday lecture session students met in lecture *after* participating in lab assignments for the week. Both lecture sections had identical reading assignments, and both sections had identical laboratory exercises, laboratory quizzes, midterms, term project and final examination.

An exception for the Friday lecture section was that several associated lab sessions met after the Friday lecture for the second of the two lab meetings each week. This exception covered 24 students of the total 109 students in the Friday lecture section.

Because of the scheduling situation in fall 2004, it was decided to operate CSE 131 such that the two lecture sections would receive different treatments:

- lecture material for the Monday lecture section consisted of the typical **introduction** of a unit. Lectures largely paralleled assigned readings for the unit. MATLAB problems were worked that were drawn from examples in the assigned readings.
- lecture material for the Friday lecture section consisted of **wrap–up** for a unit. Lectures focused largely on two areas: (a) demonstrating MATLAB points that beginning students are likely to misunderstand and (b) working MATLAB problems drawn from the exercise sets that students were assigned for lab sessions.

The two different lecture treatments presented an opportunity for retrospective, experimental comparison of student performance under "lecture as introduction" versus "lecture as wrap-up." The single metric selected for student performance was the total of "course points" earned by a student over the entire term (of a possible 100) plus the number of "extra credit points" earned (of a possible 4). This metric, of course, was also the basis for student term grades.

Initial characterization of datasets

For simplicity of presentation, we will refer to the dataset of term scores for the students in the *lecture before laboratory* treatment as the "lecture-before" dataset, and the dataset of

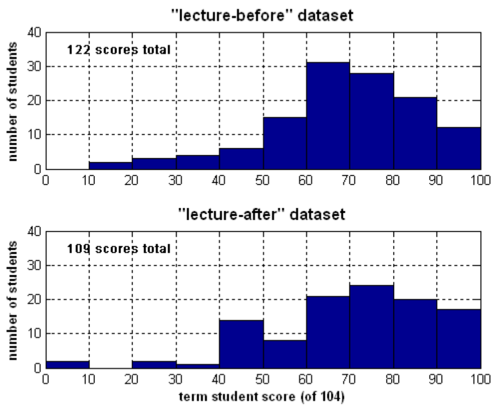

Figure 1: Histograms raw data

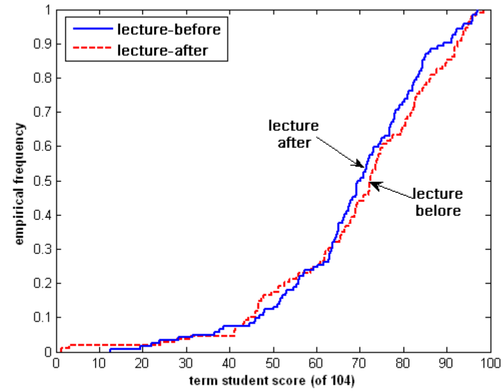

Figure 2: Cumulative frequency distributions

term scores for the students in the *lecture after laboratory* treatment as the "lecture-after" dataset.

Both lecture-before and lecture-after datasets are not statistically drawn from normal populations as revealed by application of the Jarque-Bera test. The Jarque-Bera test is a quantitative test of the difference between the observed skewness and kurtosis of a dataset and that which would be expected if the dataset were drawn from an underlying population that was normally distributed.

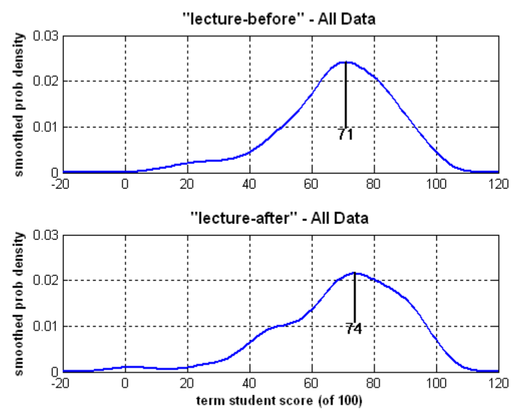

Figure 3: Smooth probability distributions

The Jarque-Bera test applied to the lecture-before dataset yields the result that the dataset is not normally distributed at the 5% confidence level (`p= 0.00000507`). The Jarque-Bera test applied to the lecture-after dataset yields the result that the dataset is not normally distributed at the 5% confidence level (`p= 0.000349`).[1]

Results of applying a two-sample Kolmogorov-Smirnov test on the lecture-before and the lecture-after datasets results in rejecting the hypothesis that the two datasets are drawn from different underlying populations at the 5% confidence level (`p= 0.50384`).

Histograms for both lecture-before and lecture-after datasets are shown in Figure 1; Figure 2 shows the experimental cumulative frequency for the two datasets; and finally, Figure 3 shows a smoothed probability distribution for the two datasets. The charts in Figure 3 were obtained with the MATLAB tool *ksdensity*, which utilizes a normal kernel function for creating the estimate. Thus the charts of Figure 3 should be used only as an exploratory tool given the earlier results that the lecture-before and lecture-after datasets were not drawn from an underlying normal population.

[1] All statistical tests reported were carried out using tools in the MATLAB Statistics Toolbox.

As noted above, the Kolmogorov-Smirnov two-sample test indicates that the lecture-before and lecture-after datasets were not drawn from statistically significantly different populations. However, there remain several points of interest in comparing the charts of Figures 1-3.

First, in Figure 2, there is a hint that the two distributions systematically differ. Note that for scores higher than a "cross over" point at a score of approximately 65, students in the lecture-after situation perform somewhat better compared to students in the lecture-before situation. 65 is the nominal (pre-curve) level for passing the course. ***Second***, also in Figure 2, note that between scores of approximately 40-60, the lecture-before students seem to systematically outperform the lecture-after students.

Thus the conclusion about Figure 2 is that there appear to be systematic differences between the lecture-before and lecture-after datasets that are dependent on the *range of scores* on which attention is focused. This in turn implies the possibility that the datasets lecture-before and lecture-after may each be drawn from more than one underlying population.

Third, focusing on the histogram in Figure 1 of the lecture-after dataset, note the secondary peak in scores in the 40-50 range. This visual observation would again be consistent with the possibility that the datasets may be drawn from multiple underlying populations.

Characterization of the complete datasets lecture-before and lecture-after yields no definitive conclusions. Application of two-sample Kolmogorov-Smirnov test indicates that there are no statistically valid differences between the samples for the lecture-before and the lecture-after datasets. The difference in the peaks for the smoothed probability functions for the two entire datasets as shown in Figure 3 (74 for lecture-before and 71 for lecture-after) must be treated as an artifact both because (a) the result from the Kolmogorov-Smirnov test indicates no statistically valid difference in the underlying populations of lecture-before and lecture-after and (b) because the smoothed probability densities shown in Figure 3 were generated with a normal kernel procedure while the Jarque-Bera tests for normalcy indicated that neither dataset could be characterized as drawn from a normal population.

However, as noted above, visual inspection of Figure 1 and Figure 2 indicates the possibility that the single dataset lecture-before and the single dataset lecture-after may each be drawn from multiple underlying populations. This is not a surprising suggestion for most instructors of large lecture session, freshman courses. From many anecdotal conversations, it is a common perception among instructors for freshman engineering courses that the total population of incoming students is drawn from multiple populations.

In the next section, we will describe our partitioning of the two complete datasets, lecture-before and lecture-after, into partitions, justification of the partitioning of the univariate datasets, and our characterization of the sub-datasets resulting from the partitions.

Partitioning the datasets; Characterization of partitioned data

To further analyze the lecture-before and lecture-after datasets, the k-means [4] non-hierarchical clustering algorithm was applied to both datasets. The k-means iterative procedure takes as inputs (a) a dataset, (b) a user-specified number of clusters, k, and (c) a metric for determining point-to-point distances between points in the dataset. The output from k-means is k groupings of the points in the dataset so as to minimize the cumulative measure of inter-point, within cluster distances.

K-means is often applied to multivariate datasets by clustering on one (or more) of the variables, and examining the result on the other variables. In our application, we have univariate datasets. K-means is still applicable provided care is exercised to ensure that the clustering is qualitatively justified by some external reason to suspect that the sample comprising the dataset is drawn from multiple populations and that the results of clustering produce well defined and distinct clusters.[5]

There are two reasons to suspect multiple underlying populations for the lecture-before and lecture-after datasets. First, as noted earlier, a common anecdotal perception among instructors in high enrollment freshman engineering courses is that class makeup consists of students drawn from multiple populations. Second, results from the section above included seemingly systematic effects in given score ranges of the experimental data.

The cluster tightness resulting from k-means clustering can be qualitatively obtained by examination of a *silhouette plot* for the clusters obtained. The silhouette value for each point in a cluster (found by application of k-means) is a measure of how close that point is to points in its own cluster compared to points in other clusters, and ranges between -1 and +1. The complete silhouette chart is a plot of all silhouette value for all clusters. A rule of thumb applied to evaluation of silhouette plots is the more points with a silhouette value of 0.8 or higher, the better the "cluster fit" to the data.

We analyzed the two datasets by clustering each into (a) two partitions and (b) three partitions. The distance metric used in the k-means clustering and in generating silhouette plots was the squared Euclidean distance, thus emphasizing the importance of the cluster "tightness." The silhouette plots for all four application of k-means is shown in Figure 4.

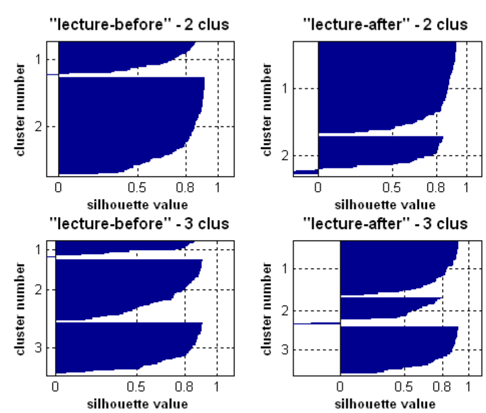

Figure 4: Silhouettes for 2 and 3 cluster solutions

Examination of the four silhouette plots in Figure 4 shows that either two or three cluster solutions produce groupings of scores that are relatively tightly organized, thus amounting to a "good clustering result." The one exception is the silhouette for the #2 cluster in the three

Table 1: Three cluster partitioning results						
	low performing cluster		*mid performing cluster*		*high performing cluster*	
	before	**after**	**before**	**after**	**before**	**after**
number	15	23	58	46	49	40
range	[12,48]	[1,54]	[50,73]	[55,77]	[74,97]	[79,99]
centroid	35	40	64	68	84	88
median	38	46	65	69	83	88
result JB test (5%)	**normal** (p=0.046)	**not normal** (p=0.0012)	**normal** (p=0.046)	**normal** (p=0.046)	**normal** (p=0.15)	**normal** (p=0.17)
result KS2 (5%)	**distributions SAME** (p=0.11)		**distributions DIFFERENT** (p=0.0032)		**distributions DIFFERENT** (p=0.025)	
result ranksum (5%)	**medians SAME** (p=0.10)		**medians DIFFERENT** (p=0.00070)		**medians DIFFERENT** (p=0.0066)	

cluster partitioning for the lecture-after dataset (lower right quadrant in Figure 4). This #2 cluster is the "low performing cluster" and would be judged of marginal tightness.

Table 2: Two cluster partitioning results				
	low performing cluster		*high performing cluster*	
	before	**after**	**before**	**after**
number	31	32	91	77
range	[12,60]	[1,62]	[62,97]	[63,99]
centroid	45	46	76	80
median	50	47	76	80
result JB test (5%)	**normal** (p=0.069)	**not normal** (p=0.00035)	**normal** (p=0.060)	**normal** (p=0.073)
result KS2 (5%)	**distributions SAME** (p=0.80)		**distributions SAME** (p=0.11)	
result ranksum (5%)	**medians SAME** (p=0.92)		**medians DIFFERENT** (p=0.04)	

Table 1 and Table 2 contain characterizations and the results of statistical tests applied to the clustered datasets resulting from k-means clustering. Table 1 shows results of three-cluster grouping; Table 2 shows results of two-cluster grouping. The first row in each table identifies a subgroup; the second row in each table identifies the dataset from which cluster scores were drawn. The remaining rows contain characterizations of resulting clustered dataset:

- **number** is the number of scores in each clustered group,
- **range** is the minimum and maximum values of scores in the clustered group,
- **centroid** is the cluster centroid of the clustered groups (from k-means),
- **median** is the median value of each clustered group,
- **result JB test** is the result of applying the Jarque-Bera test for normal distribution at the 5% confidence level to each clustered group,
- **result KS2 test** is the result of applying Kolmogorov-Smirnov two sample test for distribution differences at the 5% confidence level to paired clusters so that the result is a comparison between the lecture-before treatment and the lecture-after treatment, and
- **result ranksum** is the result of applying the ranksum test for differences between medians of two sample datasets at the 5% confidence level to paired clusters so

that the result is a comparison between the lecture-before treatment and the lecture-after treatment.

Figure 5 shows the smoothed probability distribution for the two cluster groupings; Figure 6 shows the smoothed probability distribution for the three cluster groupings. Remembering that the smoothing procedure (using the MATLAB tool *ksdensity*) utilizes a normal kernel, and using the results of the Jarque-Bera test shown in Table 1 and Table 2, the smoothed probability distribution for the *low performing cluster* in both the three cluster grouping and the two cluster grouping should be discounted and are included in the figures for completeness only.

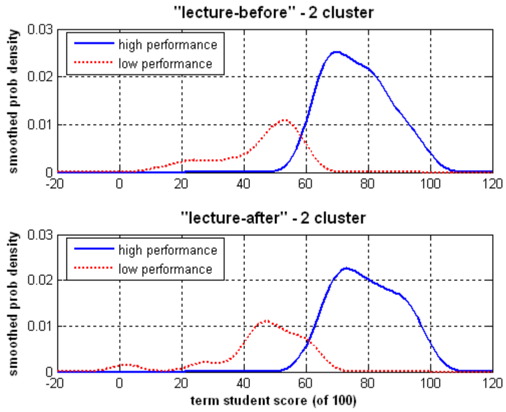

Figure 5: 2 cluster smoothed probability distributions

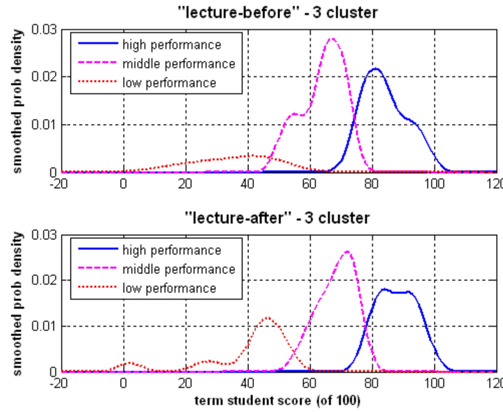

Figure 6: 3 cluster smoothed probability distributions

Summary and implication of results

There are two results from our study, one from the topic we set out to explore and one from a path we had to traverse. The focus of research reported here was to explore the effect of two modes of lecture delivery in high enrollment, gateway courses in freshman engineering: *lecture as introduction* versus *lecture as wrap up*.

Our focus was not on examining the hypothesis that high enrollment, freshman engineering classes are populated from multiple underlying populations. For the two univariate datasets of scores from the *lecture as introduction* and *lecture as wrap up* treatments, we found there was no statistically significant difference between student scores under the two differing treatments. However, as shown in Figure 2, we observed what appeared to be differences if attention was focused on results in *ranges* of scores. For example, scores above approximately 65 seemed to be better for the *lecture as wrap up* treatment. Because 65 is nominally the threshold for a passing grade in the course, we were lead to consider the possibility of multiple underlying populations of students.

We were thus lead to considering partitioning the data in each dataset into two and three cluster groups, and then analyzing the resultant subset groups, and in particular testing for differences between like groups in the *lecture as introduction* and *lecture as wrap up* datasets. Further support to justify the clustering of our univariate datasets was from the common external perception by instructors in large, freshman courses that class populations are typically drawn from multiple underlying populations of students.

The clustering was accomplished using the k-means clustering algorithm. Mild support for the internal consistency of the clustering was obtained by examination of the silhouette plots of the resultant clusters indicating that the clusters formed were relatively tight and distinct from each other.

Analysis of the sub-datasets formed by clustering indicate that no valid statistical conclusions can be drawn from the two cluster partitioning of the original datasets, as shown in Table 2. However, for the three cluster partitioning, as indicated in Table 3, scores from the mid-performing clusters and the high performing clusters were statistically significantly different. Moreover, in both cases the *lecture as wrap up* treatment produced median cluster scores that were statistically higher than the *lecture as introduction* treatment. Given the statistical validity of this result, the smoothed probability distributions shown in Figure 6 should be taken as more than artifice for the mid-performing group and the high-performing group.

The first implication from our work confirms the importance of active learning. Students in typical high enrollment, freshman engineering classes are presented with lectures that largely mirror reading assignments. Following the "standard model" of large lecture classes, students are assigned readings, come to lecture to get a second dose of the same material, then proceed to lab/recitation. Unfortunately, many students simply do not engage in this process until the time has come to "perform" – that is, many students do not read assigned material prior to lecture, but only just before they are asked to perform in lab/recitation meetings. In fact, many students simply skip the step of doing assigned readings altogether.

An "end run" on this problem of lack of student engagement in their own learning is both possible and effective as suggested by our results. Following our *lecture as wrap up* model, students go to lab/recitation *before* they attend lecture. Because they are asked to "perform" in lab/recitation, students are fully aware that they must do required readings because they will be graded on performance during the lab/recitation meeting. The lecture meeting then plays the role of clearing up misconceptions and further applying the material of the course unit via demonstration of problem solutions.

The second implication of our work is more of a question than a result: What are the factors that partition our entry level freshman students? Candidate factors include (a) student background in science and mathematics classes in high school, (b) student ability, (c) student motivation, especially because these are students making the transition from high school to university classes, (d) individual student learning styles, …

Our results are preliminary. We have built a case that *lecture as wrap up* is effective in our target course that is based on a "model-consistency argument." In our model, starting from common freshman instructor intuitions and hints from our initial analysis, we assume the existence of multiple underlying populations of students. From that assumption, we moved on to explore the consequences of the assumption by partitioning the data (using k-means), then to comparing like partitions between *lecture as wrap up* and *lecture as introduction*. This is not a definitive style of statistical argument, but rather best characterized as exploratory.

In our next steps, we intend to obtain post hoc student GPA data and use it to perform grouping of the students, then compare results of those grouping between *lecture as wrap up* and *lecture as introduction*. For freshman, GPA data is either not available at all, or is based on just one term of class work. As reported in [6], ACT results correlate to a freshman level, high enrollment computer literacy course at only about the 0.5 level. Hence university GPA and ACT scores were ruled out as external measures that could be used to partition our data, and we performed the analysis reported above based on internal partitioning. However, by Fall 2005, there will be reliable GPA data for the students whose work was analyzed in this study. We intend to obtain that GPA information and reanalyze our data accordingly.

In addition to utilizing an external factor to partition our students, in the future we look forward to performing another experiment in which we also intend to systematically measure student attitudes towards the *lecture as wrap up* delivery method. Anecdotally, and only as anecdote, we note that students seemed to participate in the *lecture as wrap up* grudgingly, and with some antagonism that tended to trail off late in the term. This initial antagonism may, if actually present, be a result of effectively forcing students to engage more fully as opposed to getting most (or all) of their surface understanding passively from lecture and with no reliance on assigned readings. In succeeding years, we intend to systematically investigate this issue of attitude.

Citations

1. Bransford, J.D., A.L. Brown, and R.R. Cocking, *How People Learn: Brain, Mind, Experience, and School*. 1999, Committee on Developments in the Science of Learning, Commission on Behavioral and Social Sciences and Education, National Academy Press: Washington, D.C.
2. Hall, S.R., et al. *Adoption of active learning in a lecture-based engineering class*. in *Frontiers in Education*. 2002. Boston, MA: IEEE.
3. Urban-Lurain, M. and D.J. Weinshank, *"I Do and I Understand:" Mastery model learning for a large non-major course.* Special Interest Group on Computer Science Education, 1999. **30**: p. 150-154.
4. Kaufman, L. and P.J. Rousseeuw, *Finding Groups in Data. An Introduction to Cluster Analysis*. 1990, New York: Wiley.
5. Jain, A., *personal communication*. 2004.
6. Urban-Lurain, M. & Weinshank, D.J., *Do non-computer science students need to program?*, Journal of Engineering Education, 90 (4) p 535-544.

Biographical Sketches

Jon Sticklen is an Associate Professor in the Department of Computer Science and Engineering at Michigan State University. He has a strong research record in knowledge-based systems. His main contributions have been in the theory and application of task specific approaches and in model-based reasoning. Dr. Sticklen has led the effort to rejuvenate the MSU College of Engineering freshman gateway course in computational tools.

Marilyn Amey is Associate Professor in the Department of Educational Administration and Chair of the Higher, Adult, and Lifelong Education Program at Michigan State University. She was part of a research team studying best practices in STEM Undergraduate Reform for SRI and NSF, and policy evaluator for an NSF Rural Systemic Reform project on math/science curriculum reform in the Navajo Nation.

Taner Eskil is a Ph.D. candidate in the Department of Computer Science and Engineering at Michigan State University. Mr. Eskil holds a M.Sc. in Mechanical Engineering and will soon complete his Ph.D. research in the area of internet agent support for electronic commerce. Mr. Eskil has been instrumental in developments in the College of Engineering freshman gateway course in computational tools.

An academic specialist in the MSU Mechanical Engineering Department, Timothy Hinds teaches undergraduate courses in machine design and statics as well as advises senior engineering student teams working on industrially sponsored capstone design projects. He also teaches a senior-level undergraduate international design project course and has taught graduate-level courses in innovation and technology management.

Mark Urban-Lurain is Director of Instructional Technology Research and Development in the Division of Science and Mathematics Education at Michigan State University. He is responsible for providing vision, direction, planning and implementation for using technology mathematics and science education and developed several introductory computer science courses for non-computer science students serving 2000 students.