ASEE's 123rd Annual • Conference & Exposition • New Orleans, LA • June 26-29, 2016

New Orleans JAZZED about Engineering Education

Paper ID #17120

Unseen Influences on Student Performance: Instructor Assessment Styles

Dr. Elif Miskioglu, Bucknell University

Elif Miskioglu is currently a Visiting Assistant Professor of Chemical Engineering at Bucknell University. She graduated from Ohio State University in 2015 with a PhD in Chemical Engineering, and is interested in student learning in engineering.

Unseen Influences on Student Performance: Instructor Assessment Styles

Abstract

Mass and energy balances is the common first course in chemical engineering (ChE) programs across the nation. This foundational course is essential for technical understanding, and thus, high concept mastery among students is desired. Highly variable student performance (in the form of grades, the widely accepted means of assessing student mastery) at a large Midwestern University suggests, however, that concept mastery is not always attained.

Learning styles are one way to describe how individuals gather and process information. The Felder-Silverman learning styles model consists of four dimensions, with two opposing preferences in each dimension that categorize individuals based on how they best process, perceive, receive, and understand information. Originally, our study focused on uncovering any correlations between student learning styles, self-efficacy, attitudes/perceptions, and performance in an undergraduate material balances course, in an effort to better understand our student population and provide a basis for curricular development. We categorized the learning styles engaged by exam problems of five instructors in their presentation and solution. While we discovered several instances where students of one learning style preference either outperformed or underperformed relative to others, we made an even more interesting observation while categorizing the learning styles exploited by exam problems of different instructors. In most cases, there was little variation between the learning styles exploited by individual instructors over the course of a semester. However, in the two instances where average student performance was statistically lower than the other three instructors' classes, the two instructors exhibited similar deviations from the other three in the types of learning styles they over- or under-exploited on exams. Further, the same two faculty had, on average, a greater number of questions and concepts on each exam than the other three instructors. An end-of-term concept inventory revealed no statistically significant difference in students' conceptual mastery between an average performance and low performance section, suggesting that performance may not be a good indicator of concept mastery in this situation.

We also observed that students in all classes consistently underperformed on questions that were categorized as "global" or "intuitive." It is arguable that at this introductory level, it is expected that students would have underdeveloped global or intuitive skills, however, if these skills do not improve over the course of their education that is cause for concern. For this reason, in future work we will be tracking students through their curricular progression in order to better understand the development of their intuitive and global skills, and assess the need for changes to the existing curriculum to foster those skills. Further, we are interested in tracking student attrition, and specifically curious as to whether students from the "underperforming" material balances classes are more likely to leave the ChE program, regardless of concept mastery. If so, this may suggest a need to develop more homogeneous course goals and means to achieve them. After multiple semesters of evaluation, we will propose a new course model that ensures a more consistent experience in this course, and hopefully a better conceptual foundation for all students.

Overview of the Work and Methods

This paper focuses on understanding how instructors (and the exams they administer) may influence student factors for success in an introductory chemical engineering course (part of the sophomore year curriculum at the institution studied). The course, commonly known as mass and energy (material) balances, is taught by two different instructors, as two separate sections, every semester. While each instructor has their own course policies, teaching philosophy, and writes their own exam problems, all instructors follow a "four exam and a final" model and use the same textbook. The exams often fall on the same day, and cover much of the same content. Thus, they provide a good basis for comparing instructors' teaching and evaluation tendencies.

After noticing large differences in raw, unadjusted (unscaled) scores between two instructors teaching material balances in the same term, we became interested in variability in student performance (as measured by grades) across all instructors of the course in the last several years. As they are the highest contributor to students' final scores, our focus is on characterization of exams administered by different faculty, and subsequent student performance on these exams.

Our hypothesis, based on initial observations from early semesters of the study, was that student performance will vary by faculty, and by the types of exam problems given by each faculty instructor. We hypothesized that students in sections with highly sequential and sensing exam problems, as well as fewer overall concepts covered on each exam, would demonstrate higher performance.

In short, we examined exams administered by six different faculty instructors (denoted Faculty A through F) to develop "faculty profiles" based on the types of exam problems administered. We characterized exam problems by the learning styles (from the Felder-Silverman model[1] of learning styles) they engaged in either presentation or solution. This characterization was done using a criterion-table developed by us, and each exam problem was categorized by multiple trained chemical engineering graduate students (specifically with B.S. degrees in chemical engineering) based on the criterion-table to ensure accurate and consistent categorization. We use the term "inherent bias" to refer to the learning styles engaged by a specific problem, and differentiate between presentation bias and solution bias.

We further evaluated these exams by the number of problems and concepts per exam. We then examined the six instructors' exams, and identified common features among instructors shown to have "low performing" classes.

Because the exam problems are written by each instructor, in Spring 2015 we also introduced a standard concept inventory[2] to measure conceptual understanding across two sections of the course. We hypothesized that differences in instructor would be highlighted by the results of a standard inventory, where neither instructor has written the questions, and students are being evaluated on a broad range of relevant course topics.

Our interest in the Felder-Silverman model of learning styles as a categorization tool for exam problem types comes from exploration into the best applications for learning styles theory in teaching. Learning styles describe how individuals receive, perceive, process, and understand information.[1,3,4] These interactions with information are essential components of problem-solving, and thus, suggest that learning styles may be a valuable lens through which to evaluate our methods for developing students as problem solvers. We used the Felder-Silverman model specifically because of its historical application in engineering, and its multidimensional nature allowing for two preferences in each of four dimensions (active/reflective, sensing/intuitive, visual/verbal, sequential/global) with subsequent strengths (strong, moderate, balanced) for each preference. This multi-dimensional model accounts for different facets of learning, and additionally emphasizes that these preferences are not fixed characteristics but merely, as they are called, preferences. Though not a specific aim of this work, we hypothesized that faculty do have learning style preferences or simply habits that unconsciously dictate their instruction, evaluation, and assessment strategies. That is, we began this study expecting to see that faculty exam problems would reveal inherent biases weighted towards certain preferences.

Statistical analysis was performed using an ANOVA followed by a Tukey-test, or Kruskal-Wallis with subsequent Steel-Dwass test as appropriate, all at a significance level of 0.05.

Results and Discussion Part 1: Learning Style Profiles across Four Semesters

In all semesters studied, students were given the Index of Learning Styles (ILS) Questionnaire[5] to evaluate their learning styles using the Felder-Silverman model. Within each dimension the class learning styles profiles have very little variation from semester to semester (Figure 1). While Figure 1 displays aggregate data for the two sections taught concurrently, both sections showed similar distributions in each semester. Our students are largely balanced across the dimensions, except for in the visual/verbal dimension where they are mostly visual learners. This does not agree entirely with previous work[1] suggesting that engineering students are active, sensing, visual, and global learners on the Felder-Silverman scale. We do note that among students that have a preference, there are more sensors and sequentials in the associated dimensions. Thus, while most students are balanced, it is important to note that there are very few intuitors, verbal learners, and global learners.

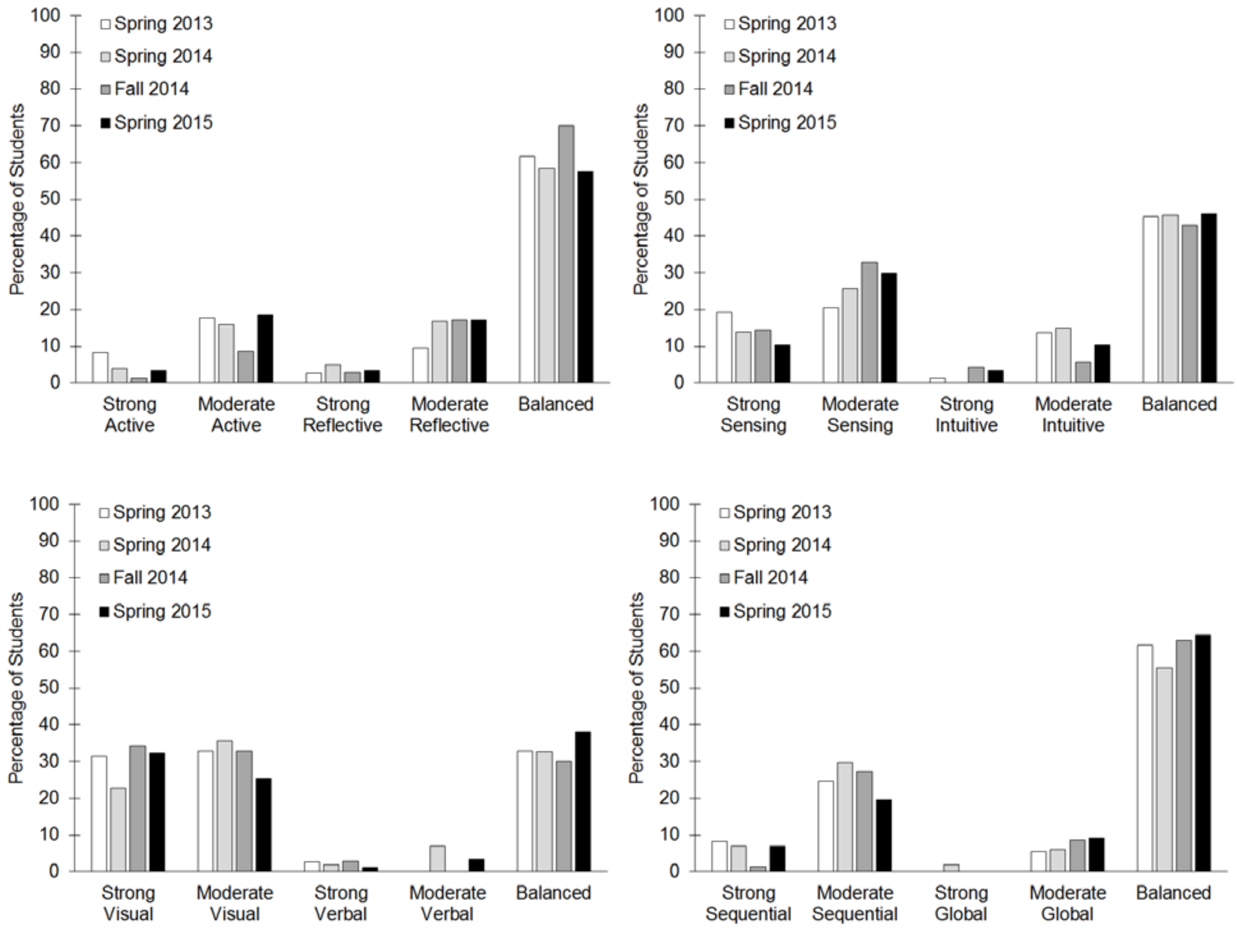

Figure 1: Learning styles profiles, by dimension, across four semesters of study. There is little variation among student population profiles throughout the duration of the study.

Results and Discussion Part 2: Variations in Teaching Style

Our attention was first drawn to instructor-driven variability in student performance when we observed in Spring of 2014 that Faculty D scaled up their raw final scores by 20% to match the course averages of Faculty B. Further, we had observed throughout the semester that Faculty B and D gave very different exams. While the content was in theory the same, Faculty B typically had two questions per exam, and they were highly sequential and sensing (numerical). Faculty D, on the other hand, had up to six problems per exam, and engaged learning style preferences across the Felder-Silverman dimensions.

With these observations, we became more interested in variability in student performance across different sections of material balances, and whether faculty with “low performing” sections shared any similar features in their exams.

Data on final grades were gathered for six faculty (aforementioned Faculty A through F) over five semesters (Spring 2013, Fall 2013, Spring 2014, Fall 2014, and Spring 2015). Faculty B, C, and D taught the course twice in this time period, whereas Faculty A, E, and F each taught the course once. There is no statistically significant difference in final grade mean between Faculty A, B, C, and F. Faculty D adjusted raw scores up at the end of term for both semesters they taught. These final adjusted grades for Faculty D show no statistical difference with Faculty A, B, C, and F. However, the unadjusted (raw, calculated) grades are significantly lower than Faculty A, B, C, and F. Faculty E's final grades also show statistical significant difference from Faculty A, B, C, and F (and adjusted Faculty D), with the highest frequency of scores being in the 70-79 range of scores (whereas most sections had the highest frequency of scores in the 80-89 range).

Without further assessment of the faculty approaches to the class (assignments, use of class time, detailed grading practices, etc.), it is difficult to draw strong conclusions as to the cause of these differences in grades. One potential contribution may be from problem presentation and solution biases. Figure 2 summarizes inherent biases in problem presentation (left) and solution (right) on exam problems written by different faculty, as a percent occurrence. It is important to note that uncategorized problems, of which Faculty E had several, are not represented in this figure. A problem went uncategorized if it did not exploit a particular learning style. Additionally, this figure does not differentiate between a "full" problem demonstrating an inherent bias versus only a portion of a problem (i.e., part (a) of a multi-part problem). Any occurrence of a presentation or solution bias on an exam is included.

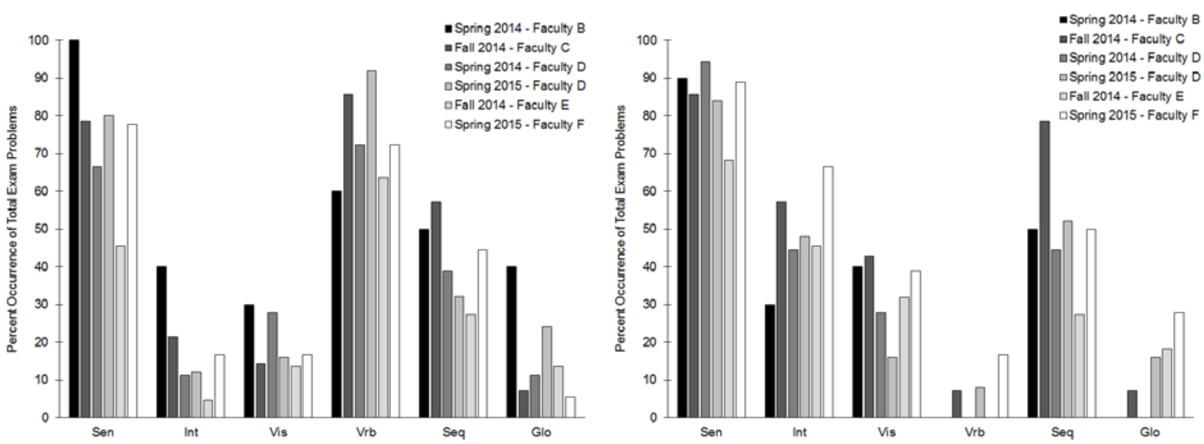

Figure 2. Summary of inherent bias by dimension for different faculty, problem presentation (left) and solution (right). X-axis shows learning style using first three letters as abbreviation (e.g., Sen = Sensing). Overall instructor biases are largely similar.

Trends across faculty appear largely similar with respect to favored problem types. Sensing, verbal, and sequential were the predominant presentation biases, and sensing, intuitive, and sequential the predominant solution biases. Contrary to our expectations, the initial observations noted from Spring 2014 between Faculty B and Faculty D's exams are not made explicit in this figure.

We do see, however, that Faculty E and D, the faculty with the lowest raw class scores, also had the lowest occurrence of sequentially presented problems and the lowest occurrence of visual solutions. In other work we noted that students are not inclined to draw a visual representation of a problem unless explicitly asked to do so. This may explain, in part, the lower performance in Faculty D and E's sections. On these exams, students may have proceeded through their problem-solving without drawing a process diagram, and thus may have been hindered in their solution process. It is important to note that while we had access to final grades from Faculty A, we did not have access to exams. We would expect that Faculty A's exams would be more similar to Faculty B, C, and F, since their final course grades were more similar to these faculty.

Considering the number of problems on each exam (Table 1, 'N/A' indicates exams that were not available to investigators), both Faculty D and E had a greater number of questions per exam (between 4 and 6) than Faculty B and F (between 2 and 4), which we believe may imply that more concepts were covered on each exam. However, Faculty C also had between 4 and 6 questions on their exams, so this trend is not consistent. A detailed categorization of concepts on exam problems, where each exam is evaluated separately, indicate that Faculty D, E, and F covered more concepts on exams than their peers (data not shown). Further, Faculty D, E, and F used more engineering-specific terminology. While neither the number of problems nor number of concepts trends were unique to Faculty D and E, along with other factors influencing teaching style (e.g., problem biases), they may have been a contributing factor to variation in final course grades across sections.

Table 1. Number of problems administered on each exam. 'N/A' indicates exams that were not available to the investigators. Faculty D and E average more questions per exam.

	Exam 1	Exam 2	Exam 3	Exam 4	Final Exam
Spring 2014					
Faculty B	3	2	3	2	N/A
Faculty D	4	6	4	4	N/A
Fall 2014					
Faculty C	4	3	3	4	N/A
Faculty E	5	6	6	5	N/A
Spring 2015					
Faculty D	4	6	5	4	6
Faculty F	4	3	3	4	4

As previously mentioned, in Spring 2015 we administered a standard concept inventory developed by Ngoathai et al.[2] to the two sections (taught by Faculty D and F). Recognizing that grades are not the best indicator of performance, and it is hard to compare across sections with no common exams, we wanted to measure student concept mastery and note whether there were any differences in concept mastery between the sections.

Students were given access to the concept inventory online during the final week of classes, and it was made available to both sections for the same five-day period. Students had two hours to submit their answers, helping to ensure that it was completed in a single sitting but also so not to impose a potentially stressful time constraint. Further, the inventory was not a graded portion of the course, but rather used as a review tool before the final exam, mitigating any incentive to cheat. Students were given access to the inventory regardless of consent to use their data, and all students were highly encouraged to give an honest effort as preparation for exams. The faculty mentioned that they would guide upcoming exam review based on inventory results as an added incentive to give a good faith effort, regardless of study participation. Thus, despite the scores not being part of student grades, we believe the concept inventory provides a good representation of student performance on conceptual tasks.

Figure 3 summarizes the concept inventory scores in Faculty D and F's sections. Performance between sections was not statistically different (though differences may seem apparent in Figure 4, as Faculty F's students outperform Faculty D's on several problems and vice versa, these were not statistically significant differences in performance and were thus not analyzed further). It is important to note that the student that had only three correct answers from Faculty D's section also spent only two minutes completing the inventory. Thus, their score should be considered an outlier. Thirty-three and thirty-seven students in Faculty D and F's, respectively, sections completed the inventory. We believe all but the aforementioned student gave a full effort, based on time on task as recorded by the learning management system (expected time was 30 minutes, based on a 20 minute average among faculty and graduate students taking the inventory). In both sections, 9 and 11 out of 20 were the most frequent scores. We do observe that the results from students in Faculty F's section more closely follows a normal distribution than the results of Faculty D's class, but again, these differences are not statistically significant.

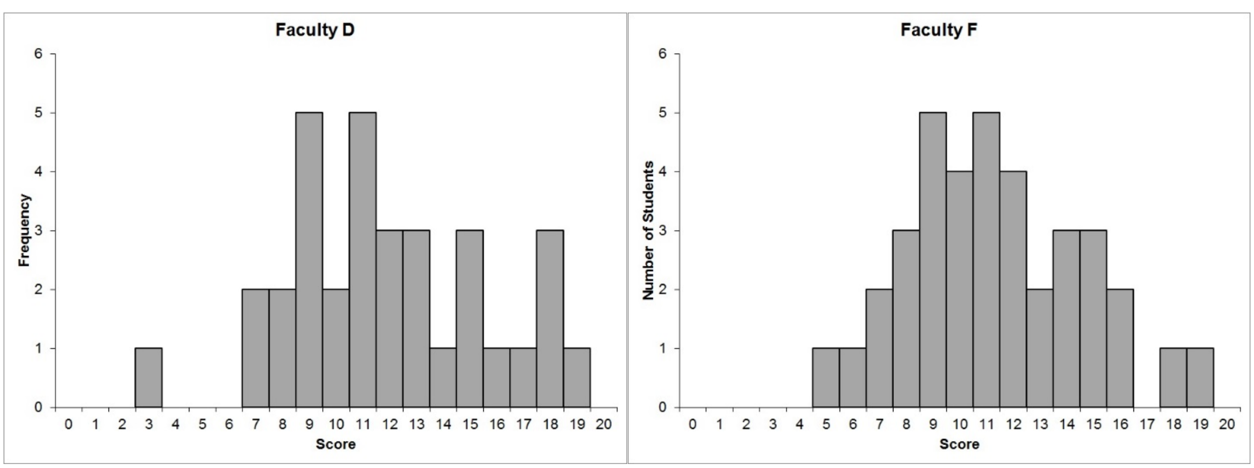

Figure 3. Distribution of concept inventory scores for students in Faculty D and F's sections. No statistically significant difference in scores was observed between faculty sections.

Figure 4 displays the fraction of students that answered each question correctly. For convenience, a horizontal line is drawn at 0.5 on the y-axis to highlight problems where fewer than half of the students provided the correct answer. There are five such questions (making up one quarter of the inventory - specifically, questions 14, 15, 17, 18, and 20), where fewer than half of the students in each section answered the problem correctly. There was also low success on questions 7 and 12 in Faculty D's section, while questions 10 and 13 had lower scores for Faculty F's section.

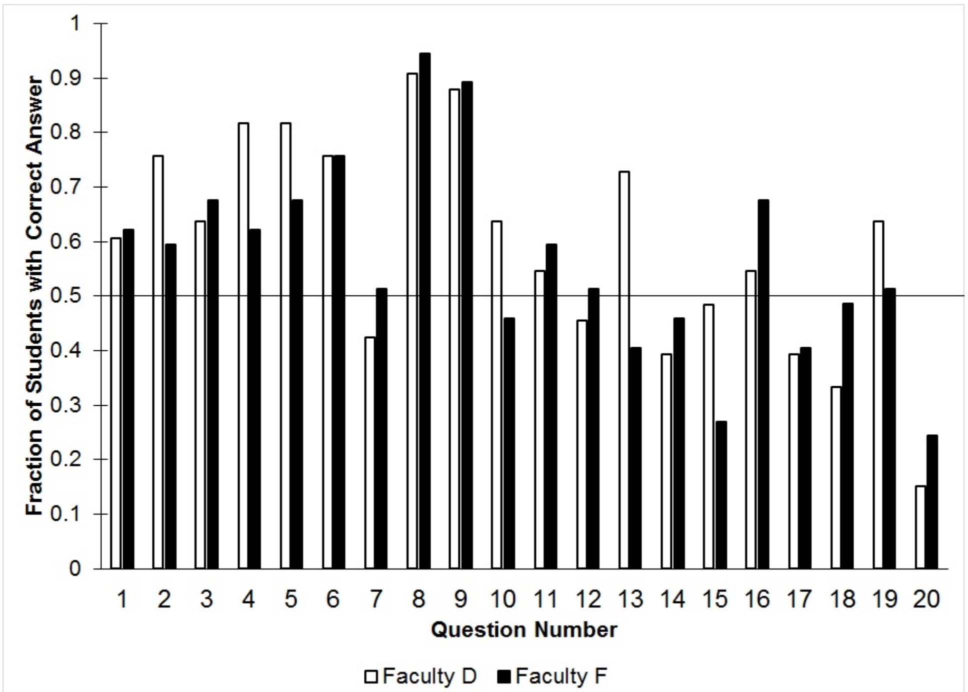

Figure 4. Comparison of student performance on each concept inventory question between Faculty D and Faculty F's section. No statistically significant difference in performance between sections was observed. Horizontal line is drawn at 0.5 to easily show problems on which 50% of the students answered incorrectly.

Because of the similar concept inventory results between sections, rather than focus our remaining analysis on what faculty outcomes may have resulted in the student performance we observed, we focused on what these "low performance" problems had in common in an effort to better understand what overarching gaps may exist in our course. We turned again to using our criterion table to categorize the problems (by learning style) in which student success was less than 50%. Without going into excessive detail about each problem, these problems were largely intuitive and global in their solution bias. While this is not surprising, as these types of holistic and theoretical problems are generally considered harder to grasp, fluency with these types of global and intuitive problems is important in developing versatile problem-solving skills. Keeping in mind that these students are at just the beginning of their chemical engineering studies, this begets a new question—does this result indicate a deficiency in our program, or rather, does it merely reflect the state of academic development our study population is in? We can only truly answer this question through a longitudinal cohort study of the student population, and assessing whether they develop and demonstrate greater intuitive and global skills as they matriculate through the program.

Conclusions and Future Work

We see that there is very little variation in ILS profiles for students entering our program from term to term, with students being largely active/reflective balanced, sensing/intuitive balanced, visual, and sequential/global balanced. Sensing and sequential are the more common preferences for students who did report a distinct preference in the associated dimension. Our results do not strictly agree with literature reports on engineering students being active, sensing, visual, and global, which may imply that our population is different or that learning styles in the greater population may be changing.

We observe variability in final grade distribution by faculty instructor, with Faculty D and E both giving lower scores (statistically significant) than their colleagues. Inspection of exams administered by these faculty suggest that this may be a result of fewer sequential problems (in other work, we have seen students report preference for sequential problems) and fewer visual solutions on their exams. Further, these faculty tended to have a greater number of problems per exam, to have a greater number of concepts covered per exam, and use more engineering terminology as compared to most other faculty. Each of these factors may have played a role in student performance.

Administration of a standard concept inventory showed little variation between students from two different sections (taught by Faculty D and F). However, some general trends within the population were observed. Problems that were answered correctly by fewer than half of the students were intuitive or global. This may suggest that these are weak areas for the student population. Whether this is cause for concern or a reflection their relatively early stage in their studies is yet to be determined.

At this juncture, there are many avenues of further investigation that may better illuminate faculty differences and encourage student success. We will continue to investigate the variability in instruction in the material balances course with more detailed assessment of course content and methods of delivery. Further characterization of instructor teaching is required, and a better understanding of evaluative processes used by each instructor (e.g., quantitative measures for comparing grading tendencies to control for “harder” or “easier” grading). Further, we intend to pursue a longitudinal cohort study of the Spring 2015 and subsequent students. We will follow them through our program in an effort to identify factors that may influence their future success, as well as track their cognitive development through the program. This will allow us to identify whether there are educational gaps that we need to address, or if observations such as intuitive and global problem weaknesses in our sophomores are merely a natural part of their intellectual development.

References

1 Felder, R. M. & Silverman, L. K. Learning and Teaching Styles in Engineering Education. *Engr Education* **78**, 674-681 (1988).

2 Ngoathai, Y. & Davis, M. C. Implementation and analysis of a Chemical Engineering Fundamentals Concept Inventory (CEFCI). *Education for Chemical Engineers* **7**, e32-e40 (2012).

3 Felder, R. M. & Spurlin, J. Applications, Reliability and Validity of the Index of Learning Styles. *Int. J. Engng Ed* **21**, 103-112 (2005).

4 Litzinger, T. A., Lee, S. H., Wise, J. C. & Felder, R. M. A Psychometric Study of the Index of Learning Styles©. *Journal of Engineering Education* **96**, 309-319, doi:10.1002/j.2168-9830.2007.tb00941.x (2007).

5 Soloman, B. A. & Felder, R. M. Index of Learning Styles Questionnaire.