

Statistical Evaluation of a “Boot Camp” Course for Preparing Students for Success in a FORTRAN Programming Course

Hiroshi Fujinoki¹, Kenneth J. Christensen¹, and Dewey Rundus¹

Abstract

Evaluation of new educational methods traditionally uses randomly selected treatment and control groups. However, if it is believed that a new method is effective then it may be desirable to allow anyone who wants to participate to do so (i.e., to not withhold the new “treatment” from possibly deserving students). This complicates evaluation of the effectiveness of the new method. We evaluate the effectiveness of a weekend “boot camp” overview course used to prepare students for a three credit hour FORTRAN programming course. Boot camp students did remarkably better in the overall course, but is this because of what was learned in boot camp or is it because the students entering boot camp were “better” to begin with? Using collected survey data and final course grades from both boot camp and non-boot camp students (totaling 256 students of which 73 enrolled in boot camp over two semesters), we attempt to answer this question. We break down the students into groups by age, gender, year, major, transfer status, experience, GPA, course load, and external load. Hypothesis testing shows that boot camp helped (or appears to have helped) most groups of students. We show that students younger than the class average received a greater benefit from the boot camp than did older students. Boot camp also benefits underclassman (freshman and sophomores) more than upperclassman. These observations suggest that the boot camp is most effective for students in the early stages of college, motivating us to encourage new students to take the boot camp. With these methods, we hope that the cause of improvement in final class grade for some student groups can be convincingly stated as being due to boot camp. Certainly, the stated results should be enough to convince any student to enroll in boot camp; the one weekend pain is small compared to the likely (but, not guaranteed) gain.

1. Introduction

When a new educational method, for example a new method of delivering a course, is introduced, it is important to evaluate its efficacy before a wide-scale deployment. Ideally, such evaluation should entail the use of randomly selected control and “treatment” groups. The treatment group receives the new educational method and the control group does not. A simple statistical evaluation of a measurable outcome can then determine if the treatment is beneficial. However, if it is believed that the new method is beneficial, it is difficult to rationalize denying some arbitrary students the treatment. Instead, it would be desirable to allow any student who wishes to participate to receive the treatment and the remaining students then form the control group. This method of self-selection of a treatment group, however, introduces an unknown into the evaluation. It is possible the self-selecting students all have a common property, or factor, that would bias this group to do well on a measurable outcome whether or not they receive the treatment. For example, if a weekend preparatory class is offered then it may be that only “better students” will participate and that these better students would have scored higher on the final outcome (e.g., final course grade) with, or without, the weekend preparatory class.

¹ Department of Computer Science and Engineering, College of Engineering, University of South Florida, Tampa, FL 33620.

In the College of Engineering at the University of South Florida, a three credit hour *Computer Tools for Engineers* (EGN 2210) course [1] is offered to all engineering students. This course is the traditional FORTRAN programming course and is, for many students, the first engineering course that they take. Enrollment in EGN 2210 ranges from 120 to 260 students. The course contains lecture, written exams, hands-on laboratories, and hands-on quizzes. The attrition level in this course is close to one-third. This high attrition level is not desirable. Two of the authors of this paper (Christensen and Rundus) along with a graduate student (Davis) developed a weekend “boot camp” course [2] to serve as an overview of the key material in EGN 2210. This weekend boot camp was voluntary in enrollment keeping with the principle that any student who wants help should be able to get it. For two semesters, campers and non-campers were tracked and their final class grades measured as the outcome. The final outcome shows that campers had a final grade that was 4 points higher than non-campers on a 100-point scale (for students that finished the course). Further, it was found that campers were 2.7 times less likely to drop the course. Are these results because of boot camp, in spite of boot camp, or simply a statement of the type of student who would enroll in boot camp in the first place? This paper describes the statistical methods to find answers to these questions.

The organization of the rest of the paper is as follows. Section 2 describes existing work in statistical methods for evaluation of self-selected groups. Section 3 describes the methods used to evaluate boot camp. Section 4 contains boot camp evaluation results. Section 5 is a summary and description of future work.

2. Existing Work

The experiment method is used to measure the effects of a given treatment by observing the responses of the treated group relative to the other group where the treatment is not assigned. Rosenbaum describes the experiment method as, “In an experiment, the assignment of treatments to subjects is controlled by the experimenter, who ensures that subjects receiving different treatments are comparable” [3]. However, it is not always possible to ensure that the subjects receiving different treatments are comparable. For example, to measure the effects of a poisonous substance to the human body, the poisonous substance cannot be given to the sample subjects to measure the effects. Cochran introduced a statistics method, called observational study, to measure the effect of a treatment where the assignment of treatments to subjects cannot be performed in a controlled manner [4].

In the experiment method, since the sample subjects are randomly separated to either the treated or controlled group, the inherent bias in the initial sample subjects will be removed after the randomized separation. Therefore, the hypothesis tests after an experiment correctly calculates the significance in the difference from the observed responses. The key difference between the experiment method and observational study is in the random separation of the sample subjects into the treated and controlled groups. Randomized selection removes self-selectivity in the sample subjects. The self-selectivity is a bias in the initial sample subjects for taking or not taking a treatment. For example, if the effects of a new treatment for cancer are to be measured and if the treatment cannot be assigned to patients since the harm from the treatment is still unknown, the effects of the treatment may be tested only on the patients with terminal cancer for whom all the ordinary treatments did not work. In this example, measuring the average survival rate of the treatment group may not correctly measure the actual effect of the treatment, since the assigned patients are already in worse health than are non-terminal patients.

In the observational study, some procedures to remove the self-selectivity must be performed before conclusions are drawn. To remove potential bias in the self-selectivity, the samples that did not take a treatment will be matched to the subjects that took the treatment. The selection of the non-treatment subjects should be done based on the variables that are believed to have influence to the effect of a treatment. This variable is called a covariate and it is simply a characteristic measured prior to the observation [3]. For the previous example of measuring the effect of a poisonous substance, possible covariates are gender, age, and tendency to chemical allergy. Thus, to compensate for the possible bias in

the observed samples, the corresponding non-treatment samples will be matched to the samples that took a treatment so that they have the same properties in the observed covariates. Then the observed results are compared to those from the non-treatment group. In case more than a single sample has exactly matching covariates, those samples can be grouped together. The resulting non-overlapping groups are called strata and the operation to classify each sample into a stratum is called exact stratification [3]. Stratification where samples have similar, but not exact, covariate matches is called inexact matching [5].

Existing work using the observational study method includes Morton et al.'s work regarding the lead level in the blood of children [6]. Morton et al. measured the level of lead in the blood of children whose parent(s) worked in a factory where lead was used. Morton chose the age of the children and the neighborhood of the children's residence as the covariates in the study. First, for each child whose parent was exposed to lead in his or her job, children who had the same covariates were matched. Then, the levels of lead in the blood were compared between the child whose parent was exposed to lead and those whose parents were not. Morton found that the children with a parent exposed to lead had a higher level of lead in their blood by 15 µg/dl compared to the children whose parents were not. Other existing work using the observational study method can be found in [7, 8, 9, 10 and 11].

Before final conclusions are made for the effects of a treatment, the significance of the difference between the responses from the treated and controlled groups should be confirmed. A Student's T-test [12] can be performed to insure that the difference in means of two groups is significant. Significance can be at a 95%, 90%, or other level. A single-tailed test is appropriate where it is to be determined if it is significant that the mean of a treatment group is greater than that of a control group. A two-tailed test applies when it is to be determined if only the mean of the treatment group is different from the control group.

3. The Matched Sampling Method and One-Tailed T-Test

We wish to determine if it is significant that the mean of an outcome of a group 1 is greater than the mean of a group 2. That is, we wish to determine if $\bar{X}_1 > \bar{X}_2$ for a 90% and 95% confidence level. We use a one-tailed T-test for samples (the groups) taken from populations with unequal variances (heteroscedastic). A test that assumes unequal population variances is a more conservative test than a homoscedastic test which assumes equal population variances. To compute t_{obs} we compute the sample means and variances of groups 1 and 2 as \bar{X}_1 , \bar{X}_2 , s_1^2 , and s_2^2 , respectively. For n_1 and n_2 samples in groups 1 and 2, respectively, we compute,

$$t_{obs} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (1)$$

where degrees of freedom, $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$. To determine t_{crit} we use the total degrees of freedom, $df = (n_1 - 1) + (n_2 - 1)$. The value of t_{crit} is obtained from a table of one-tailed T score values for the desired confidence level (e.g., 95%). If $t_{obs} > t_{crit}$ then we can say that $\bar{X}_1 > \bar{X}_2$ is statistically significant.

Matched sampling is applied to the evaluation of self-selected groups as follows. First, covariates are identified to subgroup the larger groups. For example, gender and age are covariates for human samples that may affect an outcome (e.g., for a medical treatment). The covariate samples are then separated into matched samples based on a separator (e.g., greater than mean and less than mean for a continuous-valued covariate). All samples are then compared using hypothesis testing. The effects of the covariates on the final outcome are estimated from the results of the hypothesis testing.

4. Evaluation of the Boot Camp

During the two semesters that boot camp was evaluated, a total of 256 students participated in the study. In Fall 1999, EGN 2210 was taught by one of the authors (Christensen) and in Spring 2000 the course was taught by a trio of graduate students. For both semesters, boot camp was taught by the same graduate student (Davis [2]). Of the total 256 students, 73 voluntarily enrolled in boot camp (in Fall 1999, 32% enrolled in boot camp and in Spring 2000, 25% enrolled). For Fall 2000, Table 1 shows the summary statistics for boot camp. From Table 1 it can be seen that campers were 2.7 times less likely to drop the course (since there were 183 non-campers of which 20 people dropped, while there were 73 campers of which only 3 students dropped). For the final class grade for Fall 1999 (on a scale of 100), campers averaged 80.8 and non-campers 76.7 (this is for students who did not drop the course). An 80.8 is a "B" final grade, a 76.7 is a "C". Thus, it appears that boot camp had a (very) positive effect on the final outcome for Fall 1999. The final outcome for Spring 2000 is very close to the same for campers and non-campers. This section evaluates this outcome in terms of variables that are believed to have influence on the effect of boot camp.

Table 1 - Summary statistics for boot camp

	Fall 1999	Spring 2000
Number of student enrolled in EGN 2210	121	135
Number of students enrolled in boot camp	39	34
Number of students that dropped EGN 2210	11	12
- Number of non campers that dropped	9	11
- Number of campers that dropped	2	1
Average final grade (on a scale of 100)	76.7	79.9
- Average final grade for non campers	75.3	78.9
- Average final grade for campers	80.8	81.3

All EGN 2210 students were asked to complete a nine-question survey about their background. Table 2 shows the questions and possible answers for the survey. For major the possible responses were chemical, civil, computer science, computer engineering, electrical, industrial, and mechanical. The nine questions in the survey form the covariates in this study. Of the nine covariates, gender and school transfer are binary covariates. School year, intended major, and programming experience are the multivariate covariates where more than two responses exist. For school year, students who responded with either freshman or sophomore were classified together while those who responded with either junior or senior were classified together so that the students were separated into two groups. The same approach was used for previous programming experience where the possible response from the students was either one of the four different degrees of experience. The "never" and "little" were categorized together while "some" and "much" were categorized together. For the intended major, students were separated in such a way that those who responded with computer science or computer engineering were in a group and all the others were in another group. The rest of the covariates, age, GPA, course load, and external load are continuous covariates where responses from students do not take discrete values. For the four continuous-valued covariates, the mean was used to group the students and each student was classified into one of the two groups based on the mean of the covariate samples.

Table 2 - Summary of the nine covariates

Questions	Possible responses
Age	Actual age in years
Gender	Male or female
Year	Freshman, sophomore, junior, or senior
Major	Name of major
Transfer	Yes or no
Experience	Never, little, some, or much
GPA	Cumulative GPA
Load	Hours of course work this semester
External load	Hours of external activities this semester

After the students were classified into groups by the covariates, they were further separated by those who took boot camp and those who did not. For example, for covariate “age”, each student was first grouped into either the group of students who were younger than the mean (less than average) or the group of students who were older than the mean (greater than average). Each student in the two groups was further grouped into camper or non-camper. For each of the four subgroups, the final grade in EGN 2210 was used as the observed outcome. Using (1) for 95% and 90% confidence intervals, it was tested if the (possibly) higher mean final grade of campers versus non-campers was significant. Table 3 shows the results for the non-camper subgroups and Table 4 shows the results for camper subgroups.

Table 3 - Results for non-camper subgroups

Group	Mean final grade	Significance
Group 1 - age is greater than average	78.5	No significance
Group 2 - age is lower than average	77.9	
Group 1 - gender is male	76.7	No significance
Group 2 - gender is female	77.4	
Group 1 - freshman and sophomores	76.0	No significance
Group 2 - juniors and seniors	79.3	
Group 1 - computer science or engineering	80.1	No significance
Group 2 - non computer major	77.0	
Group 1 - transfer status is “yes”	77.7	No significance
Group 2 - transfer status is “no”	79.3	
Group 1 - experience is “never” or “little”	77.7	90% significant that $\bar{X}_2 > \bar{X}_1$
Group 2 - experience is “some” or “much”	81.7	
Group 1 - GPA is greater than average	81.1	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 2 - GPA is lower than average	74.6	
Group 1 - load is greater than average	79.4	No significance
Group 2 - load is lower than average	76.7	
Group 1 - external load is greater than average	75.8	95% and 90% significant that $\bar{X}_2 > \bar{X}_1$
Group 2 - external load is lower than average	79.4	

Table 4 - Results for camper subgroups

Group	Mean final grade	Significance
Group 1 - age is greater than average Group 2 - age is lower than average	79.0 83.0	No significance
Group 1 - gender is male Group 2 - gender is female	81.6 79.7	No significance
Group 1 - freshman and sophomores Group 2 - juniors and seniors	83.1 80.9	No significance
Group 1 - computer science or engineering Group 2 - non computer major	81.4 82.4	No significance
Group 1 - transfer status is "yes" Group 2 - transfer status is "no"	78.8 81.5	No significance
Group 1 - experience is "never" or "little" Group 2 - experience is "some" or "much"	81.1 82.6	No significance
Group 1 - GPA is greater than average Group 2 - GPA is lower than average	88.1 75.2	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - load is greater than average Group 2 - load is lower than average	85.3 76.9	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - external load is greater than average Group 2 - external load is lower than average	83.6 81.9	90% significant that $\bar{X}_1 > \bar{X}_2$

Table 5 shows the results of the matched sampling method comparing campers and the non-campers for the nine factors. It can be seen that campers performed better than non-campers with 95% confidence for seven factors and with 90% confidence for one factor. The only factor that did not make a significant difference is transfer status. From the results for the seven factors where significance level is more than 95%, the following observations are made. For age, students younger than the class average appear to benefit more from boot camp than did older students. The same is true for freshman and sophomores compared to juniors and seniors. Male students appear to gain more benefit from the boot camp than do female students. Boot camp also appears to have greater benefit for student with a high GPA. One of the possible reasons for this is that these better students may have better study skills (and can thus get more out of boot camp). Finally, for those students with high course or external load, boot camp appears to be beneficial. With 90% confidence, it is shown that the boot camp appears to contribute more to the students with little programming experience than to those with much experience. A significant difference was observed between non-CS campers and non-campers, while there was no significant difference observed between CS campers and non-campers. The external and course workload show similar results. Significant difference was observed for those with heavier load between campers and non-campers, while no significant difference was observed for students with less than average load between campers and non-campers.

Table 5 - Results for camper versus non-camper

Group	Mean final grade	Significance
Group 1 - camper - age is greater than average Group 2 - non-camper - age is greater than average	79.0 78.5	No significance
Group 1 - camper - age is less than average Group 2 - non-camper - age is lower than average	83.0 77.9	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - gender is female Group 2 - non-camper - gender is female	79.7 77.4	No significance
Group 1 - camper - gender is male Group 2 - non-camper - gender is male	81.6 76.7	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - freshman/sophomore Group 2 - non-camper - freshman/sophomore	83.1 76.0	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - junior/senior Group 2 - non-camper - junior/senior	80.9 79.3	No significance
Group 1 - camper - computer science major Group 2 - non-camper - computer science major	81.4 80.1	No significance
Group 1 - camper - Non CS major Group 2 - non-camper - Non CS major	82.4 77.0	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - transfer status is "no" Group 2 - non-camper - transfer status is "no"	81.5 79.3	No significance
Group 1 - camper - transfer status is "yes" Group 2 - non-camper - transfer status is "yes"	78.8 77.7	No significance
Group 1 - camper - "never" or "little" experience Group 2 - non-camper - "never" or "little" experience	81.1 77.7	90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - "some" or "much" experience Group 2 - non-camper - "some" or "much" experience	82.6 81.7	No significance
Group 1 - camper - GPA is lower than average Group 2 - non-camper - GPA is lower than average	75.2 74.6	No significance
Group 1 - camper - GPA is greater than average Group 2 - non-camper - GPA is greater than average	88.1 81.1	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - load is lower than average Group 2 - non-camper - load is lower than average	76.9 76.7	No significance
Group 1 - camper - load is greater Group 2 - non-camper - load is greater	85.3 79.4	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$
Group 1 - camper - external load is lower Group 2 - non-camper - external load is lower	81.9 79.4	No significance
Group 1 - camper - external load is greater Group 2 - non-camper - external load is greater	83.6 75.8	95% and 90% significant that $\bar{X}_1 > \bar{X}_2$

As a “visual proof” of one of the more significant observations, Figure 1 shows a graph of age versus final outcome for campers and non-campers. The numbers above the line are the count of the camper students at an age group, while the numbers below are those for non-camper students. It can be seen that younger students who took boot camp performed better (on the final outcome). This cannot be seen for the older students.

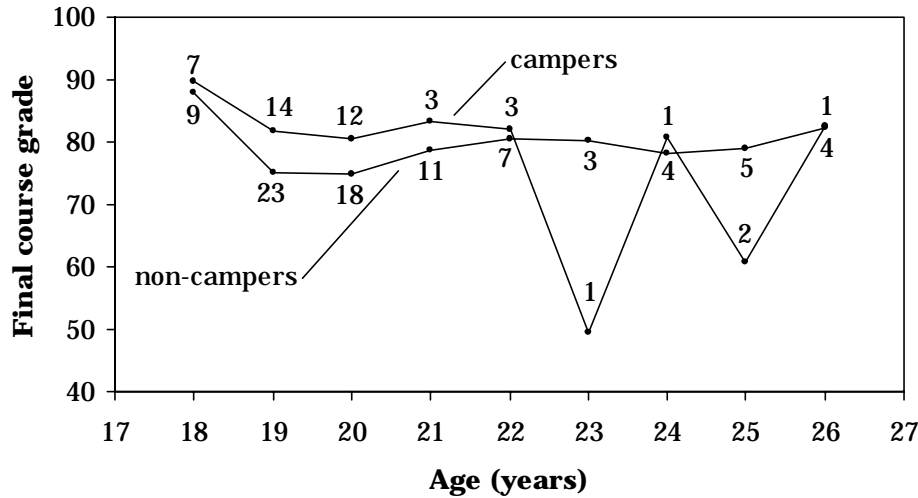


Figure 1 - Distribution of campers and non-campers for “age” covariate

5. Summary and Future Work

In this work, the matched sampling method was applied to evaluate the efficacy of a new educational method where self-selection in the students is very likely. Evaluation of self-selected groups requires different methods than evaluation of randomly selected treatment and control groups. We demonstrated that a one weekend intensive “boot camp” could have statistically significant effects on final course grades. From the results of the matched sampling method, we found that campers have a statistically significant higher final course grade than non-campers for many sub-groups. Campers younger than average and students classified as freshman and sophomore had the largest increase in final grade. We also observed that campers were 2.7 times less likely to drop the course than non-campers. The results from the matched sampling method for GPA suggest the need for more attention to students with lower GPA since the results show that the better students (those with higher GPAs) apparently benefited more from boot camper than students with a lower GPA. This suggests that future boot camps should contain training for better study skills and habits.

List of References

1. K. Christensen and D. Rundus (1998), “A First Course in Computing for Engineers,” *Proceedings of the 1998 ASEE Southeastern Section Meeting*, pp. 247 - 255.
2. K. Christensen, D. Davis, and D. Rundus (2000), “A ‘Boot Camp’ for Preparing Freshman Students for Success in a First Course in Computing,” *Proceedings of the 2000 ASEE Southeastern Section Meeting*.
3. P. R. Rosenbaum (1995), *Observational Studies*, Springer-Verlag, New York.

4. W. G. Cochran (1965), "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society, Series A*, vol. 128, pp. 134-155.
5. P. R. Rausenbaum and D. B. Rubin (1985), "The Bias Due to Incomplete Matching," *Biometrics*, vol. 41, no. 1, pp. 103-116.
6. D. Morton, A. Saah, S. Silverg, W. Owens M. Roberts, and M. Saah (1982), "Lead Absorption in Children of Employees in a Lead Related Industry," *American Journal of Epidemiology*, no. 115, pp. 549-555.
7. R. Doll and A. Hill (1966), "Mortality of British Doctors in Relation to Smoking: Observations on Coronary Thrombosis. In Epidemiological Approaches to the Study of Cancer and Other Chronic Disease," *U.S. National Cancer Institute Monograph 19, Washington D.C.: US Department of Health, Education, and Welfare*, pp. 205-268.
8. J. K. Kihlberg and S. J. Robinson (1968), "Seatbelt Use and Injury Patterns in Automobile Accidents," *Cornell Aeronautical Laboratory Report*, no. VJ-1823-R30.
9. W. G. Cochran (1968), "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, vol. 24, no. 2, pp. 295-313.
10. W. G. Cochran and D. B. Rubin (1973), "Controlling Bias in Observational Studies: A Review," *Sankhya, Series A*, vol. 35, part 2, pp. 417-446.
11. D. B. Rubin (1973), "Matching to Remove Bias in Observational Studies," *Biometrics*, vol. 29, no. 1, pp. 159-183.
12. E. Caulcott, *Significance Test* (1973), Routledge and Kegan Paul, Boston.

Ken Christensen

Ken Christensen received his Ph.D. in Electrical and Computer Engineering from North Carolina State University in 1991. He is currently an Associate Professor at the University of South Florida. His research and teaching interest is in performance evaluation of computer networks. He has over thirty conference and journal publications and eleven U.S. patents, all in the areas of computer networks and performance modeling. He was awarded a USF 1996/1997 Outstanding Undergraduate Teaching Award and a 1998 State of Florida Teaching Incentive Program award. In 1998 and 1999 he was awarded a NASA/ASEE summer faculty fellowship at Kennedy Space Center. In 1999 he was awarded a CAREER grant from the NSF. Ken is a licensed Professional Engineer in the state of Florida, a member of ACM and ASEE, and a senior member of IEEE. His homepage is at <http://www.csee.usf.edu/~christen>.

Hiroshi Fujinoki

Hiroshi Fujinoki is currently a Ph.D. student in the Department of Computer Science at the University of South Florida. His research interest is routing in computer networks. His recent contributions include developing new multicast routing algorithms and a protocol that supports real-time traffic in an internet. He received his Master's degree in Computer Science from the Illinois Institute of Technology in 1994. Hiroshi is a member of ACM and ASEE. His homepage is at <http://www.csee.usf.edu/~fujihir>.

Dewey Rundus

Dewey Rundus received his Ph.D. in Experimental Psychology from Stanford University in 1971 and his M.S. in Computer Engineering from the University of South Florida in 1985. He is currently an Associate Professor and Associate Chair in the Department of Computer Science and Engineering at the University of South Florida. His research interests focus on Human / Computer Interaction and System Usability. His concern for undergraduate education has earned him a University of South Florida Outstanding Teaching Award and a State of Florida Teaching Incentive Program Award. His homepage is at <http://www.csee.usf.edu/~rundus>.