Chapter 19
Dr. Fox Rocks: Using Data-mining Techniques to Examine Student Ratings of Instruction

Morgan C. Wang, Charles D. Dziuban, Ida J. Cook, and Patsy D. Moskal

Few traditions in higher education evoke more controversy, ambivalence, criticism, and, at the same time, support than student evaluation of instruction (SEI). Ostensibly, results from these end-of-course survey instruments serve two main functions: they provide instructors with formative input for improving their teaching, and they serve as the basis for summative profiles of professors' effectiveness through the eyes of their students. In the academy, instructor evaluations also can play out in the high-stakes environments of tenure, promotion, and merit salary increases, making this information particularly important to the professional lives of faculty members. At the research level, the volume of the literature for student ratings impresses even the most casual observer with well over 2,000 studies referenced in the Education Resources Information Center (ERIC) alone (Centra, 2003) and an untold number of additional studies published in educational, psychological, psychometric, and discipline-related journals.

There have been numerous attempts at summarizing this work (Algozzine et al., 2004; Gump, 2007; Marsh & Roche, 1997; Pounder, 2007; Wachtel, 1998). Student ratings gained such notoriety that in November 1997 the *American Psychologist* devoted an entire issue to the topic (Greenwald, 1997). The issue included student ratings articles focusing on stability and reliability, validity, dimensionality, usefulness for improving teaching and learning, and sensitivity to biasing factors, such as the *Dr. Fox* phenomenon that describes eliciting high student ratings with strategies that reflect little or no relationship to effective teaching practice (Ware & Williams, 1975; Williams & Ware, 1976, 1977).

Because of the persisting interest in student ratings, a comprehensive assortment of measurement and psychometric techniques serve as analysis models for assessing these data. Latent-trait approaches incorporated factor and component analysis in an attempt to resolve the dimensionality issues associated with these responses (Bangert, 2006; Clayson, 1999; Cohen, 2005; Feldman, 1976; Lannutti & Strauman, 2006; Marsh & Roche, 1997; Smith & Anderson, 2005). Some investigators developed hypothesis-based dimensionality studies using

---

M.C. Wang, C.D. Dziuban, I.J. Cook, and P.D. Moskal
University of Central Florida

confirmatory and hierarchical factor models; others used methods such as cluster (Ginns & Ellis, 2007) and smallest space analysis (Cohen) to define teaching profiles for effective instructors (Abrami & D'Apollonia, 1991; Apodaca & Grad, 2005; Ginns, Prosser, & Barrie, 2007). Elegant reliability studies (Chang & Hocevar, 2000) using generalizability theory resolved ratings into variance components for students, instructors, course level, items, and actions trying to account for the fact that students are nested within instructors. Other investigators incorporated classical test theory (Cook, Gelula, Dupras, & Schwartz, 2007; Lannutti & Strauman; Ustünlüoglu, 2007; Wilson, 2006).

Causal and predictive approaches applied methods such as path analysis and structural equation modeling (Chang, 2000; Ginns et al., 2007; Greenwald & Gilmore, 1997; Renaud & Murray, 2005; Rinderman & Schofield, 2001; Shevlin, Banyard, Davies, & Griffiths, 2000) that augmented more traditional regression and correlational analysis (Cohen, 2005; Davidovitch & Soen, 2006; Eiszler, 2002; Nasser & Fresko, 2006; Read, Rama, & Raghunandan, 2001; Renaud & Murray; Sheehan & DuPrey, 1999; Stapleton & Murkison, 2001). A large body of research featured hypothesis-testing models such as analysis of variance (Crumbley, Henry, & Kratchman, 2001; Maurer, 2006; Renaud & Murray; Riniolo, Johnson, Sherman, & Misso, 2006; Smith & Anderson, 2005) and chi-square contingency analysis (Howell & Symbaluk, 2001). In addition, an important approach to SEI involves deductive analysis typified by studies that incorporate criticism techniques to clarify the role of student ratings in the instructional process (Gump, 2007; Kolitch & Dean, 1999; Oliver & Sautter, 2005; Pounder, 2007). Any attempt to summarize this body of research converges on defining robust elements that underlie students' conceptions of instruction in higher education.

19.1 Student Ratings in the World of Web 2.0

During the recent decade, the emerging Internet—and in particular the concept of Web 2.0 (see http://www.oreilly.com)—impacted students' evaluations of their instructors. This phenomenon is interacting with a generation of young people on campus who have been alternatively termed millennials, the net generation, the digital generation, and generation Y, among others. Their learning and technology characteristics are described as operating at twitch speed (miniscule response time), using parallel processing for information intake, preferring information in graphic rather than textual form, using their digital, personal, and mobile technologies to stay continually connected, preferring active rather than passive learning scenarios, incorporating play into their working lifestyles, embracing learning through virtual environments, and seeing technology as fun rather than a challenge (interested readers may see http://www.marcprensky.com).

For them, the Web 2.0—with its sharing, communicating, blogging, text messaging, social networking, group writing though wikis, and interactive social opportunities—is a seamless and continuous communication medium. These developments present a learning model far different from one-directional, teacher-to-student techniques that served as the prototype for most SEI research of the past decades. Today's students experience education though online and blended courses (partly face-to-face and partly online) and extending devices, such as podcasts, chat rooms, and worldwide virtual collaborative groups.

These trends have implications for students and their instructors. One example of emerging issues is the Web site http://www.ratemyprofessors.com where students formed a worldwide community to share their perceptions about their instructors' teaching abilities. Further, they share their impressions on social networking tools, such as Facebook (http://www.facebook.com) and MySpace™ (http://www.myspace.com), or post videos of their instructors in the act of teaching on YouTube (http://www.youtube.com). On many campuses students rate their professors online rather than using the paper-and-pencil scansheets of old. Students respond, not only to their face-to-face courses, but evaluate any number of technology-mediated classes in which they might be involved.

These emerging trends make it even more important to explore elements that underpin effective teaching in the eyes of students. In order to do this, the authors explored the use of data-mining techniques to develop rule-based models that best predict what students consider excellent and poor teaching in the academy.

19.2 Data for the Present Study

The University of Central Florida (UCF) administers an end-of-course student evaluation instrument. The *Student Perception of Instruction* (SPI) form is a 16-item, Likert-type device that students use to rate their instructors (e.g., excellent, very good, good, fair, poor). Respondents have the opportunity to provide written comments about the instructor, and considerable demographic information (course level, college, department, and instructor) can be obtained from the instrument because the class and date are recorded on the form. After classes end, instructors receive the original forms with student comments and a summary of course-rating responses. Presently, many students have an online response option as well.

The instrument comprises two separately designed item sets. A university-wide committee developed the first group of eight questions, and the Florida Board of Regents provided the second set of items that were common to Florida State University System institutions. However, this distinction of item sets is not evident on the instrument. Instructors may customize the form by adding items to the preset 16-item form. No other student demographic information is collected (e.g., anticipated grade). Table 19.1 provides the items of UCF's student rating instrument.

The UCF Faculty Senate authorized a study of the results from the instrument to explore its validity for assessing alternative instructional modes. Operationally, this study sought to determine which of a number of independent variables (demographic and rating response) would predict student response to an overall rating item for their instructor.

Table 19.1 Student perception of instruction items for the University of Central Florida

Source	Questions
Administration	1. Feedback concerning your performance in this course was:
	2. The instructor's interest in your learning was:
	3. Use of class time was:
	4. The instructor's overall organization of the course was:
	5. Continuity from one class meeting to the next was:
	6. The pace of the course was:
	7. The instructor's assessment of your progress in the course was:
	8. The texts and supplemental learning materials used in the course were:
Board of regents	9. Description of course objectives and assignments:
	10. Communication of ideas and information:
	11. Expression of expectations for performance:
	12. Availability to assist students in or outside of class:
	13. Respect and concern for students:
	14. Stimulation of interest in the course:
	15. Facilitation of learning:
	16. Overall assessment of instructor:

19.3 Data Collection and Ethics Protocol

The investigators assembled a dataset containing all student ratings of instructors for the 5 academic years beginning 1996/97 through 2000/01. The file contained 588,575 student records with responses to the 16 items and corresponding demographic information. The investigators reformatted the file so that it comprised only the responses to the 16 items (five levels) and indicators of course level (lower undergraduate, upper undergraduate, graduate), college (Arts and Sciences, Business Administration, Education, Engineering and Computer Science, Health and Public Affairs), and the academic year. No further identifying information was available in the analysis file. Throughout the study, the investigators preserved department and instructor anonymity. Therefore, this study investigated the independent measures, college, course levels, academic year, and items 1 through 15 (Table 19.1) on the SPI instrument for their ability to predict overall rating of the instructor (item 16).

19.3.1 About the Analysis

In order to explore these data, the authors incorporated decision trees (Breiman, Friedman, Olshen, & Stone, 1984), a data-mining technique that identified classification rules for an instructor receiving an excellent, very good, good, fair, or poor overall rating. Justification for the authors' approach is presented below.

First, decision trees are readily applicable to large datasets such as this. To deal with missing values, the user does not have to impute values because decision trees have built-in mechanisms, such as floating-category approaches implemented by Enterprise Miner™ (SAS Institute, 2008) and the surrogate method in classification and regression trees (CART, Breiman et al., 1984). For datasets such as this one, there are many missing values; and imputation is a very difficult, time-consuming task. Second, decision trees are among the most efficient methods for studying problems of this nature. For example, a logistic regression method cannot efficiently handle all variables under consideration. There are 18 independent variables involved here; 1 variable has three levels, the other 17 have five levels. This means the logistic regression model must incorporate 68 dummy variables and 2,278 two-way interactions. Even with today's computers, this is very difficult. On the other hand, the decision-tree approach can perform this analysis very efficiently since it needs fewer computer resources (e.g., computing time and memory) even if the investigator considers higher-order interactions. Third, decision trees constitute an appropriate method for studying this problem because many of the variables are ordinal in their scaling. Although we can assign numerical values to each category, assignment of values to each category is not unique. However, decision trees use the ordinal component of the variables to derive a solution analysis. Fourth, the rules found in decision trees have an *if–then* structure that is readily comprehendible. For example, one rule derived in the analysis found that students who selected the excellent category in both *Facilitation of learning* and *Communication of ideas* had a 96% chance of selecting the excellent category for the *Overall satisfaction* item as well (Table 19.2). Fifth, the quality of these rules can be assessed with percentages of accurate classification or odds ratios that can be easily understood. The final analysis procedure produces tree-like rule structures that predict outcomes. Customarily, researchers test the quality of the rules on a dataset independent of the one on which they were developed.

Table 19.2 Decision rules that lead to an overall instructor rating of "excellent"

	Rating					
Question	E	VG	G	F	P	Excellent (*p*)
Rule 1 (*n* = 46,805)						
Facilitation of learning	•					0.96
Communication of ideas and information	•					
Rule 2 (*n* = 3,462)						
Facilitation of learning	•					
Communication of ideas and information		•				0.85
Organization of the course	•					
Assessment of student progress	•	•				
Rule 3 (*n* = 6,215)						
Facilitation of learning				•		0.78
Communication of ideas and learning	•	•				
Organization of the course	•					
Instructor interest in your learning	•					

19.3.2 The Model-building Procedure for Predicting Overall Instructor Rating

For this study, the investigators used the CART method (Breiman et al., 1984), executed with SAS Enterprise Miner (SAS Institute, 2008). Because of its strong variance-sharing tendencies with the other variables, the dependent measure for the analysis was the rating on the item *Overall rating of the instructor*, with the previously mentioned indicator variables (college, course level, academic year, and the remaining 15 questions) on the instrument. Tree-based methods are recursive, bisecting data into disjoint subgroups called terminate nodes or leaves. CART analysis incorporates three stages: data splitting, pruning methods, and homogeneous assessment.

Data splitting into two (binary) subsets at each stage is the first feature of the model. For example, all students who selected the excellent category for *Facilitation of learning* were classified into a single category; all other students were classified into another subset. After splitting, the data in the subsets become more and more homogeneous. The tree continues to split the data until the numbers in each subset are either very small (i.e., say the number of observations is less than 100) or all observations in a subset belong to one category (e.g., all observations in a subset have the same rating). Typically, this *growing-the-tree* stage results in far too many terminate nodes for the model to be useful. The extreme case occurs when the number of terminate nodes equals the number of observations. Such models are uninformative because they produce very few rules with explanatory power.

The CART procedure solves this problem by using pruning methods that reduce the dimensionality of the system. In practice, CART splits the data into two pieces: the first dataset grows the tree, and the second prunes the tree, thereby validating the model. In practice, CART methods reduce the original tree into a nested set of subtrees. Although homogeneousness based on the training dataset can always be improved, it is not necessarily true in the validation set. Typically, because the validation data are not used in the growing process, they give an honest estimate of the best tree size.

The final stage of the analysis involves assessing homogeneousness in growing and pruning the tree. One way to accomplish this is to compute the misclassification rates. For example, a rule that produces a 0.95 probability that an instructor will receive an excellent rating has an associated error of 5.0%.

An important feature of this approach involves a performance assessment of the finally developed model—accomplished by the application of rules that have been developed and validated initially to an independently collected dataset. In this case, the data from the 1996 through 1998 academic years were developmental while additional data for the 1999/2000 and 2000/01 years provided the basis for model performance assessment. Accordingly, the model development used 424,498 observations, and the performance assessment used 164,077 independent records.

19.3.3 Consequences of Using Decision Trees

Although decision-tree techniques are effective for analyzing datasets such as this, the reader should be warned of consequences of the procedure. First, decision trees only use ranks to handle both ordinal and interval variables. At times, this might lead to lost distribution information about some variables—although the use of ranks does not create any information loss in this analysis. Second, decision-tree algorithms will combine categories if a given category variable has excessive partitions. For example, most decision-tree algorithms will combine mailing code into several groups before applying a split search. This feature, however, was not problematic in this study because no categorical variable used in this study had more than ten categories. Third, the most serious weakness of decision trees is that the results can be unstable because the technique is data-driven and small variations can lead to substantially different, final solutions. Techniques such as boosting (Schapire, 1990) and bagging (Breiman, 1996) provide some remedy to the instability of tree methodology. However, these treatments make interpretation of the rules much less intuitive, countermanding the fact that ease of interpretation is one of the most important advantages of decision-tree modeling. Therefore, we did not incorporate these techniques; instead, we used a logistic regression to confirm that the resulting rules exhibit strong validity.

19.3.4 The Rules for an "Excellent" Instructor Rating

The CART method developed three rules that predicted a high probability that an instructor would receive an overall rating of excellent while three other rules led to a poor rating. All six rules only used other questions on the SPI instrument and eliminated college membership, course level, and academic year. The final solution incorporated some combination of *Facilitation of learning*, *Communication of ideas and information*, *Overall organization of the course*, *Assessment of student progress*, *Instructor was interested in your learning*, and *Instructor showed respect and concern for students*. Table 19.2 displays the three rules that led to an overall excellent instructor rating.

Rule 1 indicates that if an instructor received an excellent rating on *Facilitation of learning* and *Communication of ideas and information* then the probability of receiving an excellent overall rating is 0.96, irrespective of college, course level, academic year, or responses to any remaining questions on the rating form. Since 41.8% of the instructors in the dataset received an excellent overall rating, the odds ratio for this rule is 2.29, indicating that instructors that conform to Rule 1 are 2.29 times as likely to get an excellent overall rating than a randomly chosen instructor.

The pattern for Rule 2 also signals instructors that are good candidates for an excellent overall rating (0.85). These individuals receive excellent for *Facilitation of learning*, very good for *Communication of ideas and information*, excellent for

Organization of the course, and excellent or very good for *Assessment of student progress*. The odds ratio associated with this pattern of responses is 2.03, indicating that these instructors are slightly over twice as likely to receive an excellent overall rating as one drawn at random.

The third rule also leads to a high probability (0.78) of an instructor being viewed excellent overall. This rule blends *Facilitation of learning* (excellent), *Communication of ideas and information* (excellent or very good), and *Organization of the course* (excellent) with an additional question: *Instructor was interested in your learning* (excellent). The odds ratio for this rule was 1.87. Of the 56,482 students whose ratings conformed to either Rules 1, 2, or 3, the largest percentage (82.9%) represented Rule 1, followed by Rule 3 (11.0%) and Rule 2 (6.1%).

19.3.5 The Rules for a "Poor" Instructor Rating

Three informative nodes produced substantially high probabilities that an instructor would receive an overall poor rating. There was a high correspondence among the questions of the SPI form that predicted an overall rating of excellent and an overall rating of poor. Once again, the poor rules used only other questions on the rating instrument and eliminated college, course level, and academic year. The poor rules replaced the question *The instructor was interested in your learning* that appeared in the excellent rules with *The instructor showed respect and concern for students*. Table 19.3 depicts the outcomes associated with the three rules.

Rule 4 illustrates if an instructor receives a fair or poor on the question *Facilitation of learning* and a poor on both *Communication of ideas and information* and *Instructor is interested in your learning* then the probability of an overall

Table 19.3 Decision rules that lead to an overall instructor rating of "poor"

	Rating					
Question	E	VG	G	F	P	Poor (p)
Rule 4 (n = 1,821)						
Facilitation of learning				•	•	
Communication of ideas and information					•	0.83
Instructor interested in your learning					•	
Rule 5 (n = 1,135)						
Facilitation of learning				•	•	
Communication of ideas and information					•	0.58
Organization of the course					•	
Rule 6 (n = 532)						
Facilitation of learning				•	•	
Communication of ideas and learning			•	•		0.54
Assessment of student progress					•	
Respect and concern for students					•	

rating of poor is 0.83. Because the percentage of instructors receiving an overall rating of poor in the dataset is 1.9%, the odds ratio for this rule is extremely high (43.6). This means that students classified in this category are significantly more likely to designate poor as the instructor's overall rating. However, the odds ratio of 43.6 might overestimate the magnitude of this likelihood.

Rule 5 states if an instructor receives a fair or poor on *Facilitation of learning* and a poor on both *Communication of ideas and information* and *Overall organization of the course*, then the probability of an overall rating of poor is 0.58. Although the probability of a poor rating with this combination of responses seems somewhat lower than the previous rule, one should note that the odds ratio associated with this rule is 30.3. This means this rule still has a significantly higher likelihood of giving the instructor an overall poor rating than a student randomly selected from the university.

Rule 6 indicates if an instructor's rating for *Facilitation of learning* is fair or poor, *Communication of ideas and information* is good or fair, and *Assessment of student progress* and *Instructor shows respect and concern for students* are poor, then the probability of an overall rating is 0.54 with an associated odds ratio of 32.3. The probability of an instructor receiving an overall rating of fair or poor for Rule 4 = 0.99, for Rule 5 = 0.97, and for Rule 6 = 0.96.

19.3.6 Model Validity

The investigators used three approaches to validating the decision-tree model—two logical and one statistical. The logical approaches involved harvesting all instructors across the university that conformed to the excellent and poor decision rules and examining the degree to which the rules leveled college differences. Table 19.4 presents the results of that procedure for excellent rules (academic years 1999/2000/2001). The unadjusted column depicts the percentages of overall excellent instructor ratings by college in the absence of the rules.

Ratings ranged from a high of 53.79% for Education to a low of 36.33% for Business Administration. The columns under *Adjusted for rule* portray the results when instructors across colleges are selected according to their compliance with the rules. In this case, the differences virtually disappear. Rule 1 produces overall excellent instructor ratings in the colleges, ranging from a high of 97.12% (Education) to a low of 95.03% (Business Administration). Rule 2 adjusts the excellent ratings from a high of 86.23% in Education to a low of 83.07% in Business Administration. Rule 3 produces a high of 80.05% in Arts and Sciences to a low of 74.00% in Health and Public Affairs. Table 19.4 demonstrates that college differences equalize around the proportions specified by each rule when instructor ratings conform to the rules that lead to a high probability of excellent.

Table 19.5 shows the impact of the poor rules on instructor ratings across the colleges. Again, the unadjusted column indicates the percentages of instructors that received an overall rating of poor, not taking into account the rules. Those

Table 19.4 Percentage of instructors receiving "excellent" overall ratings by college unadjusted and adjusted by rules 1–3

			Adjusted for rule					
College	Unadjusted	*n*	1	*n*	2	*n*	3	*n*
Arts and Sciences	41.83	31,914	95.30	19,699	85.41	1,358	80.05	2,419
Business Administration	36.33	12,463	95.03	7,950	83.07	628	77.12	974
Education	53.79	8,819	97.12	6,634	86.23	313	75.21	458
Engineering and Computer Science	32.19	4,434	95.52	2,604	83.71	185	78.16	365
Health and Public Affairs	47.80	11,138	96.13	7,894	85.53	455	74.00	632

Table 19.5 Percentage of instructors receiving "poor" overall ratings by college unadjusted and adjusted by rules 4–6

			Adjusted for rule					
College	Unadjusted	*n*	1	*n*	2	*n*	3	*n*
Arts and Sciences	2.31	1,761	78.95	630	57.14	264	54.44	135
Business Administration	3.01	1,033	85.87	383	63.36	166	50.37	68
Education	1.68	276	89.77	79	52.27	46	50.00	21
Engineering and Computer Science	4.81	662	83.69	272	56.25	90	62.90	39
Health and Public Affairs	1.89	441	87.80	144	53.99	88	53.33	24

unadjusted percentages ranged from a high of 4.81% in Engineering and Computer Science to a low of 1.68% in Education. Viewing the ratings according to the poor rules produces dramatic changes. Those instructors who conformed to Rule 4 were overall rated poor, ranging from a high of 89.77% in Education to a low of 78.95% in Arts and Sciences. Percentages of poor ratings for instructors that conformed to the pattern of Rule 5 showed the highest percentage of poor ratings in Business Administration at 63.22% with a low value found for Education at 52.27%. Finally, Rule 6 defines instructors who were rated poor in Engineering and Computer Science at a rate of 62.90% with the lowest value found in Education at 50.00%.

The second logical validation approach involved comparing the results of the UCF study with two national initiatives on teaching excellence. A model that identified seven principles of effective undergraduate education has gained widespread acceptance as a national standard for higher education (Chickering & Gamson, 1987). These seven principles describe an instructor who encourages contacts between faculty and students, develops cooperation and reciprocity, uses active learning techniques, gives prompt feedback, respects diverse talents and ways of thinking, emphasizes time on task, and communicates high expectations. A parallel initiative, the National Study of Student Engagement (Kuh, 2001), described five benchmarks: student interaction with faculty, collaborative learning, active learning, supportive environments, and academic challenge. Table 19.6 presents the correspondences between these two initiatives and the UCF-CART study. A comparison of these initiatives shows a close correspondence with components in each system grounded in facilitation of learning, instruction interest in student learning, effective communication, a well-organized learning environment, respect for students, and effective assessment of student progress.

In order to further examine the model validity, the investigators completed two separate logistic regression analyses. Table 19.7 presents the results of the analyses for the items contributing to an overall excellent rating of the instructor. All items selected by the decision tree contributed to the equation with the Wald chi-square probabilities rounded to 0.00. The model predicted with 97.6% accuracy producing a Somer's *D* of 0.945.

The logistic regression results for those items leading to a poor overall rating selected by the decision tree are presented in Table 19.8. Once again, the analyses showed that all items in the rules produced Wald chi-square values with associated probabilities rounded to 0.00. This equation produced a predictive accuracy of 97.6% with a Somer's *D* of 0.963.

Table 19.6 A comparison of the seven principles of good practice, the National Study of Student Engagement, and UCF's rule-based items

Seven principles of good practice (Chickering & Gamson, 1987)	National Study of Student Engagement (Kuh, 2001)	UCF rule-based items (applicable rule)
Encourages contacts between faculty and students	Student interaction with faculty	Facilitation of learning (1,2,3)
Develops reciprocity and cooperation among students	Collaborative learning	Instructor interested in your learning (3)
Uses active learning techniques	Active learning	Communication of information and ideas (1,2,3)
Gives prompt feedback	Supportive environment	Well-organized course (2,3)
Respects diverse talents and ways of thinking		Respect and concern for students (6)[a]
Emphasizes time on task	Academic challenge	Assessment of student progress
Communicates high expectations		

[a]Poor rating on this item correlates with an overall rating of Poor.

Table 19.7 Logistic regression for "excellent" rule items[a]

Items	*df*	Coefficient	Wald χ^2	*p*
Intercept	1	0.895	23,385.1	0.0001
Interest	1	0.796	23,157.2	0.0001
Organization	1	0.798	23,903.1	0.0001
Assessment	1	0.847	12,454.4	0.0001
Communication	1	0.847	25,162.3	0.0001
Facilitation	1	1.092	33,328.6	0.0001

[a]Percent correctly predicted = 97.6, Somer's *D* = .963.

Table 19.8 Logistic regression for "poor" rule items[a]

Items	*df*	Coefficient	Wald χ^2	*p*
Intercept	1	0.171	349.9	0.0001
Interest	1	−0.691	6,472.9	0.0001
Organization	1	−0.919	12,650.7	0.0001
Assessment	1	−0.667	7,062.0	0.0001
Communication	1	0.924	13,598.8	0.0001
Facilitation	1	−0.961	14,867.4	0.0001

[a]Percent correctly predicted = 97.6, Somer's *D* = .963.

19.3.7 A Discussion of the Dimensions

In their book on facilitative teaching, Wittmer and Myrick (1974) provided characteristics for what students considered poor teaching. Those instructors were insensitive, cold, disinterested, authoritarian, ridiculing, arbitrary, sarcastic, demanding, punitive, and disciplinarians. The students described excellent teachers as good listeners, empathetic, caring, concerned, genuine, warm, interested, knowledgeable, trusting, friendly with a sense of humor, dynamic, and able to communicate effectively. This second list resonates with all three of the excellent rules.

Rogers (1993) described a facilitative teacher as one who creates a learning environment rather than simply transmitting knowledge. The key element in Rogers' theory of teaching emphasized the facilitator's empathetic understanding when he or she comprehended and valued a student's perceptions. Straus (1988) examined facilitation from a leadership perspective and built facets of the process that might be construed as a teaching model. His theory demonstrated seven components: sharing an inspiring vision, focusing on results process and relationship, seeking maximum possible involvement, designing pathways to action, bringing out the best in others, celebrating accomplishment, and modeling behaviors that facilitate collaboration. Not only do students respond positively to a facilitative class environment, several theories support facilitation as an effective teaching model.

The ability to communicate effectively has long been accepted as a standard for effective teaching. Our findings suggest that this ability is fundamental to an instructor being viewed positively by students. In fact, one of the terminate nodes we obtained involved only two items that led to a high probability of a poor overall rating. This happened when students rated an instructor high on interest in student learning but low on communication ability—most likely a frustrating and ambivalent situation for students.

The CART analysis suggests that students reward instructors who develop effective course organization and evaluation techniques. These components may be viewed as skills obtainable through professional development. Because of recent emerging modalities for classes (e.g., fully online, blended, Web-enhanced), course organization has gained prominent attention. In addition, instructors are under increasing pressure to make assessment of student learning an organic component in their courses. One should note, however, that organization and assessment impact an instructor's rating in the presence of facilitation and communication. By themselves, they are not strong enough to carry the instructor's rating.

Respect for students and interest in their learning weighted differently in the student evaluation process. Instructor interest in student learning contributed to an excellent rating while student perception of low instructor respect for them resulted in poor overall ratings. These results led us to conclude that a supportive class climate created by instructors is a strong motivating factor for students to view their class experience positively.

19.4 Conclusion

Classification and regression tree analysis of student rating of instruction appears to have lived up to the expectations we placed upon it. By efficiently handling missing data and multiple interactions, the procedure produced reasonably robust decision rules that identify qualities by which students characterize excellent and poor university instructors. Another advantage of the rule-based solution comes from the ability of multiple constituencies (students, faculty, administrators) to integrate results such as these into their decision-making processes. Decision-tree methodology provides compelling outcomes through probability statements, odds ratios, and misclassification assessment, thereby allowing users to judge the quality, usefulness, and opportunity costs found in rule-based outcomes. In addition to operational and specific if–then rules, another advantage derives from the tree-like structures that provide comprehensive and systematic solutions to examining student evaluation in complex systems.

This approach produces an interactive and recursive model whereby an individual (e.g., a dean of a college) might review the results of the analyses and through more extensive investigation examine the effects of additional independent variables (e.g., class size, laboratory sections, online classes) on student ratings. All these added values indicate that the decision-tree analysis, above all, is responsive to a number of elements in the emerging information society. Today's students and faculty live in a world of ambient fundability (Morville, 2005) that comprises a fast-moving society where anybody can find anyone or anything, anywhere, anytime. These findings put new pressure on faculty members to respond to students' information and communication needs while at the same time maintaining the rigor required by their disciplines. However, when decision rules portray excellent teachers as facilitative, communicative, organized, interested, and equitable, they configure a prototype learning situation far different from the traditional *paper chase* of a few years ago. One of the most efficient methods for building these profiles comes from data-mining techniques.

Seldom do new technologies replace old ones immediately but, rather, begin a complex pattern of interactions with them over a period of time. For instance, the workplace watercooler has gone digital although not completely. From the decision rules in this study, one might infer that digital networks are creating a collective intelligence in which problem solving becomes an activity of the commons where students expect a participatory learning environment. In the digital world, knowledge comes from real-life experiences (or their simulations) rather than from formal education. For many years, it was precisely that formal education to which SEI research directed its attention. Decision-tree analysis appears to work well as a flexible format for examining student responses as they evaluate a much more recursive learning environment.

Current higher education environments feature community, collaboration, and self-organization, which create learning climates that are cognitively complex, reliant on technology, and much less dependent on physical geography. Peer production becomes an important part of this new learning space, displacing many features of the academy as we have known it. The decision rules suggest that students wish to lessen the ambiguity they experience in their classes with the concomitant reduction of their ambivalent feelings toward higher education. They prefer active involvement because they participate in a highly interactive world that employs multiple learning facets.

The method of categorization and regression tress for analyzing student satisfaction with instruction is well suited to the evolving nature of higher education. By providing grounded decision rules, it avoids the difficulties encountered in nonobservable, latent-trait approaches and the prohibitive assumptions underpinning many predictive and hypothesis-testing procedures. Certainly, one must be sensitive to the fact that this method does not produce a one-time solution and that selection of the variables for inclusion in the analysis has a major impact on the results, thereby underscoring the need for context planning in such studies. In addition, investigators should be cognizant of the large number of observations required for these methods. However, the fact that the final results produce clearly interpretable rubrics permitting one to take actions on such issues as instructional design, curriculum planning, course offerings, and administrative policy bring data-mining techniques into the mainstream of higher education as a decision tool for the information age.

References

Abrami, P. C., & D'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness: Generalizability of "N = 1" research: Comment of Marsh. *Journal of Educational Psychology, 83*(3), 411–415.

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., et al. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching, 52*(4), 134–141.

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni- and multidimensional models. *Studies in Higher Education, 30*(6), 723–748.

Bangert, A. W. (2006). Identifying factors underlying the quality of online teaching effectiveness: An exploratory study. *Journal of Computing in Higher Education, 17*(2), 79–99.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 495–518.

Chang, L., & Hocevar, D. (2000). Models of generalizability theory in analyzing existing faculty evaluation data. *Applied Measurement in Education, 13*(3), 255–275.

Chang, T.-S. (2000, April). *An application of regression models with student ratings in determining course effectiveness*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED455311)

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39*(7), 3–7.

Clayson, D. E. (1999). Students' evaluation of teaching effectiveness: Some implications of stability. *Journal of Marketing Education, 21*(1), 68–75.

Cohen, E. H. (2005). Student evaluations of course and teacher: Factor analysis and SSA approaches. *Assessment & Evaluation in Higher Education, 30*(2), 123–136.

Cook, D. A., Gelula, M. H., Dupras, D. M., & Schwartz, A. (2007). Instructional methods and cognitive and learning styles in web-based learning: Report of two randomised trials. *Medical Education, 41*(9), 897–905.

Crumbley, L., Henry, B. K., & Kratchman, S. H. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education, 9*(4), 197–207.

Davidovitch, N., & Soen, D. (2006). Using students' assessments to improve instructors' quality of teaching. *Journal of Further and Higher Education, 30*(4), 351–376.

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education, 43*(4), 483–501.

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 5*(3), 243–288.

Ginns, P., & Ellis, R. (2007). Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. *The Internet and Higher Education, 10*(1), 53–64.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603–615.

Greenwald, A. G. (Ed.). (1997). Student ratings of professors [Current Issues]. *American Psychologist, 52*(11), 1182–1225.

Greenwald, A. G., & Gilmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*(11), 1209–1217.

Gump, S. E. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review. *Educational Research Quarterly, 30*(3), 55–68.

Howell, A. J., & Symbaluk, D. G. (2001). Published student ratings of instruction: Revealing and reconciling the views of students and faculty. *Journal of Educational Psychology, 93*(4), 790–796.

Kolitch, E., & Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education, 24*(1), 27–42.

Kuh, G. D. (2001). Assessing what really matters to student learning. *Change, 33*(3), 10–19.

Lannutti, P. J., & Strauman, E. C. (2006). Classroom communication: The influence of instructor self-disclosure on student evaluations. *Communication Quarterly, 54*(1), 89–99.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187–1197.

Maurer, T. W. (2006). Cognitive dissonance or revenge? Student grades and course evaluations. *Teaching of Psychology, 33*(3), 176–179.

Morville, P. (2005). *Ambient findability: What we find changes who we become*. Sebastopol, CA: O'Reilly Media.

Nasser, F., & Fresko, B. (2006). Predicting student ratings: The relationship between actual student ratings and instructors' predictions. *Assessment & Evaluation in Higher Education, 31*(1), 1–18.

Oliver, R. L., & Sautter, E. P. (2005). Using course management systems to enhance the value of student evaluations of teaching. *Journal of Education for Business, 80*(4), 231–234.

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education, 15*(2), 178–191.

Read, W. J., Rama, D. V., & Raghunandan, K. (2001). The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business, 76*(4), 189–192.

Renaud, R. D., & Murray, H. G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education, 46*(8), 929–953.

Rinderman, H., & Schofield, N. (2001). Generalizability of multidimensional student ratings of university instruction across courses and teachers. *Research in Higher Education, 42*(4), 377–399.

Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or not: Do professors perceived as physically attractive receive higher student evaluations? *Journal of General Psychology, 133*(1), 19–35.

Rogers, C. R. (1993). The interpersonal relationship in the facilitation of learning. In M. Thorpe, R. Edwards, & A. Hanson (Eds.), *Culture and processes of adult learning* (pp. 228–242). London: Routledge.

SAS Institute. (2008). SAS® Enterprise Miner™ (Version 5.3) [computer software]. Cary, NC: Author. Available from http://www.sas.com/technologies/analytics/datamining/miner/index.html

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning, 5*(2), 197–227.

Sheehan, E. P., & DuPrey, T. (1999). Student evaluations of university teaching. *Journal of Instructional Psychology, 26*(3), 188–194.

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education, 25*(4), 397–405.

Smith, G., & Anderson, K. J. (2005). Students' ratings of professors: The teaching style contingency for Latino professors. *Journal of Latinos and Education, 4*(2), 115–136.

Stapleton, R. J., & Murkison, G. (2001). Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and expected grades. *Journal of Management Education, 25*(3), 269–291.

Straus, D. A. (1988). *Facilitative leadership: Theoretical underpinnings*. Cambridge, MA: Interaction Associates.

Ustünlüoglu, E. (2007). University students' perceptions of native and non-native teachers. *Teachers and Teaching: Theory and Practice, 13*(1), 63–79.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief overview. *Assessment & Evaluation in Higher Education, 23*(2), 191–212.

Ware, J. E., Jr., & Williams, R. G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education, 50*, 149–156.

Williams, R. G., & Ware, J. E., Jr. (1976). Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. *Journal of Educational Psychology, 68*(1), 48–56.

Williams, R. G., & Ware, J. E., Jr. (1977). An extended visit with Dr. Fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecturer. *American Educational Research Journal, 14*(4), 449–457.

Wilson, J. H. (2006). Predicting student attitudes and grades from perceptions of instructors' attitudes. *Teaching of Psychology, 33*(2), 91–95.

Wittmer, R. D., & Myrick, R. D. (1974). *Facilitative teaching: Theory and practice*. Pacific Palisades, CA: Goodyear Publishing.