

**Recency Effects and Students' Course Evaluations**

Shaun McQuitty  
Michael R. Hyman  
Eric R. Pratt  
Pookie Sautter

New Mexico State University  
Department of Marketing  
College of Business Administration and Economics  
Box 30001, Dept. 5280  
Las Cruces, New Mexico 88003-8001

Shaun McQuitty is associate professor, Michael R. Hyman is professor, Eric R. Pratt is associate professor, and Pookie Sautter is associate professor in the Department of Marketing at New Mexico State University. Please send all correspondence regarding this manuscript to Shaun McQuitty (Office Phone (505) 646-1463; Fax (505) 646-1498; email: mcquitty@nmsu.edu).

© 2002 Shaun McQuitty, Michael R. Hyman, Eric R. Pratt, and Pookie Sautter

## **Recency Effects and Students' Course Evaluations**

### **Abstract**

Although recency effects have been demonstrated in many areas of academic inquiry, one area where such effects have received little scrutiny is students' course evaluations. In an empirical study that addresses this issue, ratings of the last two classes in a semester are found highly related to overall course evaluations. This intuitively consistent finding has implications for how professors should prepare for their final classes in a semester, but it also raises ethical questions about the manipulation of SET scores.

## **Recency Effects and Students' Course Evaluations**

Simpson and Siguaw (2000) write that “Probably no other fact of academic life has stirred as much controversy and research as SETs [student evaluations of teaching]” (p. 200). Because aggregate SET scores are important to college professors—due to their impact on salary, awards, promotion, and tenure decisions—Simpson and Siguaw study faculty responses to such evaluations. Among their many findings is a mention of professors' attempts to manipulate SET scores by “having an unusually great/fun class meeting/lecture/film the day before SET administration” (p. 207). That such a manipulation would succeed suggests that students have little memory of classes earlier in the semester, or at least that SET scores depend less heavily on memories of earlier classes than on memories of a final few classes.

If most students believe that class quality is invariable, then a recency effect would be undetectable. However, quality variability or heterogeneity in classroom teaching is expected because it is a form of service delivery (Berry 1980; Zeithaml, Parasuraman, and Berry 1985). A question that naturally arises, therefore, is “Do all classes during a semester affect SET scores equally?” If a recency effect unduly influences SET scores, then the last few classes in a semester will affect them disproportionately. This prediction implies that professors who want to receive the best possible course evaluations should ensure that the classes immediately prior to SET administration appeal to students.

The magnitude of the recency effect on SET scores has not been studied in an education context. Therefore, this article presents an exploratory study that addresses the question of whether or not perceptions of end-of-semester classes affect SET scores disproportionately (i.e., Does a recency effect influence SET scores?). After a brief description of this study, evidence for a recency effect bias is presented and its implications are discussed.

## **Order Effects and SET Scores**

Order effects occur when the sequence in which information is presented to people affects their responses. Two types of order effects are the recency effect and the primacy effect. The recency effect implies that the most recent information is weighted more heavily than earlier information, and the primacy effect suggests that the earliest information is weighted more heavily than later information. Order effects are well understood in some areas of study, such as social psychology and consumer behavior (e.g., Johar, Jedidi, and Jacoby 1997; Richter and Kruglanski 1998; Sinclair 1988; Steiner and Rain 1989). As a typical example, the initial and final items in a list of items are more memorable, i.e., more easily retrieved from memory, than items in the middle of the list (Routh and Frosdick 1978).

Prior research on SET scores has found evidence for a primacy effect bias (e.g., Feldman 1977), which implies that students' initial impressions of a professor or course are significantly related to end-of-semester evaluations. In this vein, Liu and Ginther (2001) report that “successfully managing social and interpersonal considerations, especially the formation and initial impressions, often substantively influences both learning outcomes and student evaluations of the learning experience” (pp. 50-51).

Although primacy effects have been studied in the context of SET scores, recency effects have been ignored in the same context. If students have relatively short memories of classes throughout a semester, then their impressions of classes immediately prior to SET administration should be more strongly related to SET scores than classes from earlier in the semester.

In the next section an empirical study of a recency effect in a pedagogical context is described. Because official SET scores are confidential, two surrogates—measures of students' satisfaction with a course and their impressions of its quality—are employed instead. The

empirical results indicate that the last two classes in a semester are significant predictors both of students' satisfaction with a course and their rating of its overall quality. Discussion of these findings and the associated implications follows.

### **Study Description and Results**

Students in two large sections of an introductory marketing class at a medium-sized university in the southwestern U.S. responded to two different questionnaires for extra course credit. The first questionnaire, which obtained from students their assessment of a class period's quality and their satisfaction with that class, was administered at the end of each class for eleven consecutive classes near the end of a semester. No questionnaire was administered on a day when an examination was given. A second questionnaire, administered during the last class in the semester, asked students about their perceptions of overall class quality and satisfaction with the class, plus a variety of other related and demographic questions. Two different professors taught the two course sections, and the final sample was comprised of 237 students.

Responses for the quality and satisfaction questions were given on a seven-point scale anchored by Excellent--Very Poor (quality question), and a nine-point scale anchored by Extremely High--Extremely Low (satisfaction question). Table 1 provides descriptive statistics for each of the eleven class satisfaction ratings and for overall satisfaction with the course (assessed during the semester's last class). Table 2 provides the same statistics for overall quality and the quality ratings for each of the eleven classes. These results show that students are fairly satisfied with each class (class means range from 5.9 to 7.2 on a nine-point scale), and the quality ratings for each class are comparable (class means range from 4.8 to 5.7 on a seven-point scale). The overall quality and satisfaction measures both used nine-point scales with the same anchors

as the individual class measures, and the mean ratings for overall satisfaction (6.68) and overall quality (6.50) are similar.

-----  
Insert Tables 1 & 2 about here  
-----

Regression analyses are used to evaluate the capacity of individual class ratings to explain variance in overall service quality and satisfaction. All regressions reported throughout the analyses employ the OLS estimation method with pairwise deletion of missing data. To begin, overall satisfaction is regressed on the satisfaction ratings for all eleven classes, producing  $R^2 = 0.54$ . Table 2 shows that ratings of the last two classes are the best predictors. Regressing overall satisfaction on the satisfaction ratings for the last two classes drops  $R^2$  from 0.54 to 0.48, so overall satisfaction strongly depends on ratings of the most recent classes. Part correlations (Part R) that remove the effects of other independent variables characterize the unique predictive value of each explanatory variable, thereby overcoming the potential effect of multicollinearity in the regression output (see Table 3). The part correlations also indicate that overall satisfaction relates more strongly to satisfaction ratings for the last two classes than to satisfaction ratings for earlier classes.

-----  
Insert Table 3 about here  
-----

Next, a similar analysis is performed for overall quality. However, when overall quality is regressed on quality ratings for each of the eleven classes ( $R^2 = 0.58$ ), ratings of the last two classes are not the best predictors (see Table 4). Only the fifth class period has a significant regression coefficient (possibly due to chance). Part correlations support this interpretation of the regression results. Nonetheless, when overall quality is regressed on quality ratings for the last

two classes, the beta coefficients for both classes are significant ( $R^2$ 's decrease from 0.58 to 0.45). Again, it appears that overall quality is strongly related to ratings of the two most recent classes, although earlier classes can also be strongly related.

-----  
Insert Table 4 about here  
-----

Seemingly, the satisfaction and quality ratings associated with recent classes can strongly impact overall satisfaction and quality ratings for courses. How do ratings of the last two classes perform when compared to traditional measures of teaching quality? Several items from the second questionnaire are utilized for this comparison, including students' ratings of the time expended on the course relative to other courses, their willingness to recommend the course, the importance they place on a college education and good grades, their current GPA and expected grade, and how much they learned and received from the course. Overall quality is regressed on these eight variables and produces  $R^2 = 0.50$ , which is only slightly better than the  $R^2$  for ratings of the last two classes as predictors of overall quality ( $R^2 = 0.45$ ). Regressing overall satisfaction on the same eight variables produces  $R^2 = 0.60$  (versus 0.48 with the satisfaction ratings for the last two classes). Of the traditional measures, willingness to recommend the course and how much students received from the class are the best predictors of overall satisfaction and overall quality. When the traditional measures are combined with ratings of the last two classes, the  $R^2$ 's climb to 0.61 (quality) and 0.71 (satisfaction), so the latter variables clearly improve explanatory power.

### **Conclusions**

The results indicate that a significant recency effect exists for university students' overall satisfaction with a course and their ratings of overall course quality. Specifically, quality and

satisfaction ratings associated with the last two classes prior to SET administration explain a practically and statistically significant portion of the variance in perceptions of overall quality and satisfaction. For overall satisfaction, the last two classes are more important than any previous classes for which measures are obtained. The same pattern is not found for overall quality, as some classes prior to the last two are more strongly related to overall quality. However, quality ratings for the last two classes are still significant predictors of overall quality.

Because university administrators weight SET scores heavily when assessing professors (Becker and Watts 1999; Simpson and Siguaw 2000), the evidence for a recency effect bias in students' evaluations clearly indicates that it behooves professors to save their best for last, i.e., to pay particular attention to prepping the last few classes in a semester. Of course, this is just one of many factors that can contribute to SET scores, and positive classroom experiences should start at the beginning of a semester and continue throughout. Nonetheless, efforts to produce especially good classes near the end of a semester can enhance students' perceptions of overall satisfaction and quality; presumably, official SET scores are similarly affected.

However compelling this study and its results, they are subject to limitations. First, because the conclusions are based on one non-random sample only, validation through replication is needed. Second, measures are available for only the final eleven classes of the semester, which precludes testing for a primacy effect. Such a test would be valuable, because several pedagogy researchers have found evidence for a primacy effect in SET scores (e.g., Feldman 1977). If quality and satisfaction data from the beginning of the semester were available, it would also be possible to compare the relative strengths of the recency and primacy effects. Third, a common-method bias may have produced an upward bias in the relationship

between items, although the study used different length scales and different descriptors for each question.

This study suggests ideas for future research. One example is that there may be particular conditions under which a recency effect on SET scores is enhanced or weakened. Additional research is needed to identify these conditions and when they apply. A second example is that these data show that satisfaction and quality are strongly related (the correlation between the two overall measures is 0.82). Measures of satisfaction typically produce skewed data (LaBarbera and Mazursky 1983; Oliver 1981; Westbrook 1980), and Peterson and Wilson (1992) conclude that

Despite the plethora of variables, perspectives, and methodologies investigated, it is not possible to succinctly identify specific causes of the striking characteristic of customer satisfaction ratings that motivated the present research [i.e., skewness]. The phenomenon is too complex for a simple answer (p.68).

Because the correlation between the overall satisfaction and quality measures is high, it might be expected that the quality measure would demonstrate a skewness similar to the satisfaction measure. However, that is not the case with these data, and a test indicates that the difference between the two distributions is significant ( $\chi^2(36) = 364.04$ ,  $P < 0.001$ ). The overall quality responses more closely approximate a normal distribution than the overall satisfaction responses. Which of these two measures provides a more valid interpretation of what students perceive? What do the overall quality and overall satisfaction measures capture that the other doesn't? Answers to these questions could be important to the debate about the relationship between quality and satisfaction in a service setting.

Finally, this research raises ethical and practical questions about the manipulation of SET scores. What happens if professors know that adjusting the way a few key classes are taught can

significantly alter SET scores? Will they manipulate the evaluation system in a way that undermines the integrity of the learning process? Alternatively, does this research provide further evidence of the need for multidimensional measures of teaching performance? The conclusion that a recency effect may affect SET scores adds new arguments to the ongoing debate about such questions.

**Table 1**  
**Satisfaction Ratings**

Class Period	Mean Rating	Std. Dev.	N	Correlation with Overall Satisfaction
1	6.117	1.523	197	0.519
2	6.263	1.512	179	0.531
3	6.143	1.702	182	0.476
4	5.924	1.509	184	0.515
5	6.369	1.612	187	0.594
6	6.000	1.543	196	0.466
7	6.184	1.751	174	0.617
8	6.264	1.658	178	0.570
9	7.204	1.324	191	0.318
10	6.179	1.720	196	0.600
11	6.567	1.502	203	0.655
Overall Satisfaction	6.675	1.411	237	1.000

**Table 2**  
**Quality Ratings**

Class Period	Mean Rating	Std. Dev.	N	Correlation with Overall Quality
1	4.949	0.991	196	0.568
2	5.028	1.016	180	0.521
3	4.973	1.105	182	0.499
4	4.755	1.071	184	0.493
5	5.032	1.107	187	0.601
6	4.776	1.023	196	0.564
7	5.017	1.223	174	0.618
8	4.836	1.244	177	0.609
9	5.660	0.970	191	0.357
10	4.970	1.202	196	0.580
11	5.157	1.053	204	0.633
Overall Quality	6.544	1.511	237	1.000

**Table 3**  
**Regression of Overall Satisfaction on Individual Class Satisfaction**

Class Period	Coefficient	Std. Error	t	P	Part R
(Constant)	2.019	0.544	3.709	0.000	
1	0.027	0.084	0.328	0.744	0.020
2	0.080	0.081	0.989	0.324	0.059
3	0.089	0.069	1.284	0.201	0.077
4	0.029	0.083	0.351	0.726	0.021
5	0.121	0.085	1.428	0.156	0.085
6	-0.038	0.076	-0.506	0.613	-0.030
7	0.103	0.081	1.275	0.205	0.076
8	0.036	0.080	0.449	0.654	0.027
9	-0.086	0.080	-1.076	0.284	-0.064
10	0.133	0.073	1.836	0.069	0.110
11	0.254	0.114	2.221	0.028	0.133

$R^2 = 0.54$ ,  $N = 237$

**Table 4**  
**Regression of Overall Quality on Individual Class Quality**

Class Period	Coefficient	Std. Error	t	P	Part R
(Constant)	-0.159	0.632	-0.251	0.802	
1	0.157	0.130	1.205	0.230	0.069
2	0.166	0.112	1.487	0.139	0.085
3	0.163	0.099	1.635	0.104	0.094
4	-0.033	0.112	-0.294	0.769	-0.017
5	0.237	0.111	2.132	0.035	0.122
6	0.076	0.130	0.586	0.559	0.034
7	0.148	0.111	1.331	0.185	0.076
8	0.192	0.105	1.839	0.068	0.105
9	0.001	0.103	0.014	0.989	0.001
10	0.125	0.107	1.166	0.246	0.067
11	0.113	0.155	0.733	0.465	0.042

$R^2 = 0.58$ ,  $N = 237$

## References

- Becker, William E. and Michael Watts (1999), "The State of Economic Education: How Departments of Economics Evaluate Teaching," *American Economic Review*, 89 (May), 344-349.
- Berry, Leonard L. (1980), "Service Marketing is Different," *Business*, 30, 24-29.
- Feldman, K. A. (1977), "Consistency and Variability among College Students in Rating Their Teachers and Courses: A Review and Analysis," *Research in Higher Education*, 6, 223-274.
- Johar, Gita Venkataramani, Kamel Jedidi, and Jacob Jacoby (1997), "A Varying-Parameter Averaging Model of On-Line Brand Evaluations," *Journal of Consumer Research*, 24 (September), 232-247.
- LaBarbera, Priscilla A. and David Mazursky (1983), "A Longitudinal Assessment of Consumer Satisfaction/Dissatisfaction: The Dynamic Aspect of the Cognitive Process," *Journal of Marketing Research*, 20 (November), 393-404.
- Liu, Yuliang and Dean Ginther (2001), "Managing Impression Formation in Computer-Mediated Communication," *Educause Quarterly*, 24 (3), 50-54.
- Oliver, Richard L. (1981), "Measurement and Evaluation of Satisfaction in Retail Settings," *Journal of Retailing*, 57 (Fall), 25-48.
- Peterson, Robert A. and William R. Wilson (1992), "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the Academy of Marketing Science*, 20 (Winter), 61-71.
- Richter, Linda and Arie W. Kruglanski (1998), "Seizing on the Latest: Motivationally Driven Recency Effects in Impression Formation," *Journal of Experimental Social Psychology*, 34 (4), 313-329.
- Routh, David A. and Richard M. Frosdick (1878), "The Basis and Implications of the Restoration of a Recency Effect in Immediate Serial Recall," *Journal of Experimental Psychology*, 30, 201-220.
- Simpson, Penny M. and Judy A. Siguaw (2000), "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response," *Journal of Marketing Education*, 22 (December), 199-213.
- Sinclair, R. C. (1988), "Mood, Categorization Breadth, and Performance Appraisal: The Effects of Order on Information Acquisition and Affective State on Halo, Accuracy, Information Retrieval, and Evaluations," *Organizational Behavior and Human Decision Processes*, 42, 22-46.
- Steiner, D. D., and J. S. Rain (1989), "Immediate and Delayed Primacy and Recency Effects in Performance Evaluation," *Journal of Applied Psychology*, 74, 136-142.

Westbrook, Robert A. (1980), "A Rating Scale for Measuring Product/Service Satisfaction," *Journal of Marketing*, 44 (Fall), 68-72.

Zeithaml, Valarie A., A. Parasuraman, and Leonard L. Berry (1985), "Problems and Strategies in Services Marketing," *Journal of Marketing*, 49 (Spring), 33-46.