Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 10 Number 6, June 2005 **ISSN 1531-7714**

Recency Effect in College Student Course Evaluations

David Dickey, University of West Florida
Carolyn Pearson, University of Arkansas at Little Rock

Universities depend on faculty evaluations to substantiate tenure and promotion decisions and it is well known that student ratings are a critical source of information for this process. Of recent concern in course evaluations is recency effect, a rating error that occurs when more weight is assigned to activities that are closer to the time of formal appraisal. Quantitative procedures used in this study revealed that students who were either trained in recency effect and kept a course diary or kept a course diary only were sensitive to this type of error on a measure of teaching skill, but not course organization. Qualitative procedures also revealed that students were aware of recency effect once it was defined.

It is well known that students' rating of instructors is perhaps the most widely used method of assessing instructor effectiveness (Centra, 1993, 2003; Chen & Hoshower, 2003; Filak & Sheldon, 2003; Ory, 2001; Spencer & Schmelkin, 2002). Many issues concerning validity of student evaluations have been examined including bias (Marsh, 1984), which was found to be minimally present based on the grade the student expected to receive (Centra, 1993, 2003), and student characteristics (e.g., grade point average, academic ability, gender, age); which some studies found did not affect student evaluations (Centra & Creech, 1976; McKeachie, 1979; Centra, 1993, 2003), and some studies found did affect student evaluations (Basow & Howe, 1987; Tang, 1999; Tatro, 1995). Since students are the recipients and raters of instruction, they can provide an important, unique, and necessary perspective on judging teacher effectiveness. Student evaluations are commonly used to make decisions regarding faculty salary, awards, and promotion and tenure; therefore, it is imperative to consider all the major variables that can influence student evaluations. Although there has been extensive research on student ratings of instruction (e.g., Marsh, 1987; Seldin, 1993), the research is only now focusing on those who partake in the evaluation process: faculty, students, and administrators (Chen & Hoshower, 2003; Filak & Sheldon, 2003; Spencer & Schmelkin, 2002). Specifically, when attempting to assess teaching quality through student evaluations, it is important to be aware of the specific behaviors exhibited by students as they evaluate faculty.

One area of student behavior that has been examined is rating errors, which are inadequacies of one sort or another in performance appraisals (Berk, 1986; Cascio, 1989; Mathis & Jackson, 1994) and the most common rating errors are halo effect, leniency or severity error (Centra, 1993), and central tendency; which have been examined extensively in the past in instructional evaluations. Halo effect occurs when a rater has the tendency to think of a person as being generally good or inferior, which would yield relatively high inter-correlations between different dimensions of performance behavior. Leniency or severity error is defined as the rater having the tendency to assign a higher or lower rating to an individual than is warranted by the demonstrated behavior. Central tendency is the

raters' unwillingness to go out on the proverbial limb in either the favorable or unfavorable direction when evaluating a person (Berk, 1986; Cascio, 1989; Mathis & Jackson, 1994). Emphasis has emerged on another rating error, defined as recency effect, which is the focus of this study. Anyone who has ever had a project fail near appraisal review time understands recency effect, or the tendency to assign more weight to activities occurring near the time of the formal review (Mohrman, Resnick-West, & Lawler, 1989). The raters' fallible information-processing capabilities cause such errors to be perceived as idiosyncratic and their perceptions are compounded by incomplete memories (Mohrman et al., 1989).

Few studies have inquired into students' general attitudes towards the evaluations such as how conscientiously they respond, how seriously they take the process, and what purposes they think they are being used for (Spencer & Schmelkin, 2002). The characteristics exhibited may have little or nothing to do with the effectiveness of a particular course or teacher when evaluating instruction (Centra, 1993) and this has been demonstrated more recently in that students are skeptical about the use of their ratings since they are unsure of whether their opinions matter or to what purpose the ratings are put to, even though they are not reluctant to do them and have no fear of bias (Spencer & Schmelkin, 2002). Although there have been cautions in the past regarding examining bias (Marsh, 1984), the intent was to focus on potential differences that affect how students perceive evaluations and reflects how the length of time the student has partaken in the educational process influences their views (Spencer & Schmelkin, 2002). Since the research seems to be shifting towards focusing on those who partake in the evaluation process, including students and the behaviors that they demonstrate (Chen & Hoshower, 2003; Filak & Sheldon, 2003; Spencer & Schmelkin, 2002), and since students are considered to be raters for the purposes of assessing instructor effectiveness, the purpose of this study was to investigate whether university students experienced recency effect as a rating error.

BACKGROUND

Recency effect, which may be a potentially important rating error, gives particular attention to one's memory or cognitive process when performance rating. Recency effect suggests that a delay between observing a person and conducting the evaluation can cause variance in ratings as a result of memory decay (Nathan & Lord, 1983). When making memory-based judgments people often rely on a general impression or a category prototype to help them access or reconstruct stored information. The error implies that halo effect will also increase as time increases (Nathan & Lord). Recency effect is also most likely to occur when appraisals are done over a long period of time. Traditionally, performance evaluations provided by students are at the end of a course, which can be four to five months long, and the literature has addressed the benefits of mid-course formative evaluations in which students have an opportunity to provide input and see possible changes (e.g., Abbott et al., 1990). Research on this issue indicates biased ratings may result under longer circumstances, especially if the information has been stored in the rater's memory according to irrelevant, over-simplistic, or otherwise faulty categories (Cascio, 1989). When academic organizations use only one instructor evaluation conducted at the end of the semester, there may be a tendency for the students to remember more about what their instructor has done just prior to the evaluation than in the prior months, although later evaluations could give students an overall perspective for assessment of the class (Spencer & Schmelkin, 2002). Recency effect is an understandable rating error because it is difficult to remember performance that may be 4 to 5 months old. Individuals being rated may also become more concerned about performance as the formal appraisal time approaches (Mathis & Jackson, 1994); and because of this, may influence or take advantage of recency effect by carrying out favorable acts just prior to the evaluation.

To increase the accuracy of information by minimizing recency effect requires performance documentation, especially by diary (Carrell , Elbert, & Hatfield, 1995; Greenberg, 1987; Maroney & Buckley, 1992; Roberts, 1995). Because performance evaluations frequently occur months after the performance is observed, raters need a means for sustaining memory (Bernardin & Beatty, 1984; Mohrman et al.). Notes or memoranda made by the rater of a performance can be filed systematically and the rater can then refer back to the documentation when appraising (Carrell et al., 1995; Cascio, 1982; Mathis & Jackson, 1994). College student evaluations need to be timed appropriately in order to ensure an effective evaluation and requiring every course to be evaluated may contribute to deterioration in data quality. Students may consider the evaluation of every course they take as excessive and can become

careless in filling out the appraisal; whereas, when students know that they do not always get an opportunity to express their opinions, they may be more careful in giving the appraisal their full attention (French-Lazovik, 1982).

PURPOSE

Based on the review presented it is evident that there are still issues concerning the validity and reliability of student evaluations of college instruction, especially pertaining to the behaviors that students demonstrate when rating the instruction they have received. It is also apparent that course evaluations are a product of information that has been encoded and stored in memory by students for later retrieval, which provides the opportunity for recency error, but may be alleviated with performance documentation. It was stated earlier that students are considered raters for the purposes of assessing instructor effectiveness; therefore, the purpose of this study was to investigate whether university students could be sensitized to recency effect as a rating error. Specifically, the study sought to determine if students were aware of recency effect and if receiving training and/or keeping a diary impacted the course evaluation. We hypothesized that students who received awareness training in recency effect and kept a course diary, or kept a course diary only, would demonstrate significantly different course organization and teaching skill scores from students who did neither, and that these same two groups would demonstrate more stable scores over the term.

METHODS

Participants

Responses were gathered from 113 students enrolled in a core technology course required by all undergraduate programs within a college of education in a major Florida university. The study utilized four course sections and participants were enrolled in various program disciplines with no specifically targeted classification or instructional area sought for this study in order to aid generalizability of the results. The four course sections were randomly selected from a pool of courses that were accessible to the researchers, but were then all taught by the same instructor in order to control that effect. Also, five additional instructors from the pool of courses and 15 randomly selected participants from the four sections examined were interviewed in this study. The instructors and students interviewed were a simple random sample taken from a spreadsheet of their names that was created at the beginning of the study, and after deletion of the instructors, would later be used for the course grades of the students.

Procedures

The procedures used in the experiment were based on qualitative and quantitative research methods. For the qualitative portion, a semi-structured interview was used to investigate if students were even aware of recency effect as a rating error. The interview involved asking a series of structured questions, and based on the response, probing more deeply using open-ended questions to obtain additional information (Gall et al., 1996). The questions were developed in advance and were field-tested by the primary author, who was the only person conducting the interviews. For the quantitative portion of the study, a modification of the Solomon 4-group design was used to achieve two purposes: to assess the effect of the two experimental conditions (training plus diary and diary only) and to determine the presence of pretest sensitization (Gall et al., 1996).

The study utilized four separate sections of the core technology course in which pre- and post-testing was conducted using the university course evaluation form during the first and last week of classes, respectively. In order to link the pre and post course evaluations, the students were asked to create a unique identifier to put on the forms and to keep the number in a place where they would remember it. The design was structured so that the experimental condition consisted of two separate classes (training plus diary and diary only, both pre-and post-tested) and the control groups consisted of two separate classes (one both pre- and post-tested and one post-tested only). The pre-tested groups were asked to fill out the form based on what they had heard about the course and the instructor; all groups were administered the same form at the end of the course as a posttest.

The first group's treatment consisted of a 15- to 20-minute training presentation and handout on rating error, which emphasized recency effect, and orientation to the university student evaluation form. In the presentation the researcher provided a definition of recency effect and examples, and then had the participants provide experiences with or examples of recency effect. The group was then instructed to keep a diary throughout the course on the instructor's

performance utilizing the university course evaluation form as their primary reference. At a minimum, the students documented weekly what went well and poorly overall and specific good or bad behaviors that were demonstrated by the instructor. The second group's treatment consisted of keeping a diary only, and instructions were the same as the first group. Participant failure to document weekly in the diary was an issue in terms of internal validity; therefore, the primary author provided a notebook for such use. Communication via email between the participants and primary author, plus in-class reminders by the instructors, was the primary means of ensuring treatment fidelity.

Instrumentation

The university course evaluation form consisted of 16 items that were divided into two sections, organization of course and teaching skill; scores were obtained for each section by summing the respective items and could range from 8 to 40, respectively. Participants responded to the items using a modified Likert-type scale that ranged from 1 (strongly disagree) to 5 (strongly agree). Course organization items inquired into such aspects as course expectations, requirements, assignments, testing procedures, grading practices, and class meetings, etc.; teaching skill items inquired into such aspects as instructor's command of the subject, communication of information, respect and concern for students, availability for assistance outside of class, etc.

Data Analysis

For the quantitative analysis, a modification to the Solomon 4-group design was made to include two treatment groups (training and diary, diary only) and a pretested control group; therefore, group comparisons were made using analysis of covariance using the pretest scores as the covariate, and the organization and teaching skill scores as dependent variables. All statistical assumptions pertinent to ANCOVA were examined prior to the main data analysis and Cohen's d was used to determine effect sizes. Interpretational analysis was used to analyze the qualitative data using a semi-structured interview. Interpretational analysis is the process of examining case study data closely in order to find constructs, themes, and patterns that can be used to describe and explain phenomenon (Gall et al., 1996).

RESULTS

Quantitative Analysis

The training/diary group contained 30 students, the diary group contained 29 students, the pretested control group contained 25 students, and the non-pretested control group contained 29 students (N = 113). There were no missing data for the analysis and the design was fairly balanced; any participant that did not have complete pretest and posttest data was dropped from the study. Random selection was used to assign the classes to the levels of the treatment, and random assignment of the participants to the classes was assumed since the sections were formed via the student registration system. The four classes all met at different times during the week and any student interaction throughout the semester was probably minimal; therefore, it was assumed that the groups had minimal influence on each other's evaluation. The distributions of the organization and teaching skill scores within the groups were normal and there were no outliers, and the homogeneity of variance assumption was met. Prior to the ANCOVA, an examination of the group and pretest interaction revealed that the trend lines were ordinal; thus, the assumption of homogeneity of regression slopes was met [$F(2,78) = .06$, $p > .05$; $F(2,78) = .52$, $p > .05$; for organization and teaching skill, respectively].

To determine if pretest sensitization occurred, the two control groups were compared on both organization ($M = 35.12$, $SD = 4.67$; $M = 38.14$, $SD = 2.23$) and teaching skill post-tests ($M = 35.28$, $SD = 4.75$; $M = 38.45$, $SD = 2.53$) for pretested and non-pretested respectively, and significant differences were detected for both variables, [$t(52) = 3.10$, $p = .003$; $t(52) = 3.12$, $p = .003$]. Examination of the posttest means for all groups did reveal that the control group that did not receive the pretest scored the highest on both variables; thus, the differences detected were probably due to pretest sensitization since the pretested groups were instructed to complete the evaluation form based on what they had previously heard about the course and instructor (See Table 2). The means and standard deviations for the four groups on the pretest and posttest course evaluation items are presented in Table 1 and gains were made on all 16 items of the course evaluation; the greatest gain was made in the pre-and post-tested control group, followed by the diary only group, and then the training and diary group.

.

Table 1 *Means and Standard Deviations of Pre- and Post- Organization and Teaching Skill*

	Training and Diary				Diary Only				Pre- and Post-test Control				Post-test only Control	
	Pre		*Post*		*Pre*		*Post*		*Pre*		*Post*		*Post*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Organization														
Expression of expectation for performance in this class	3.83	0.87	4.40	0.62	3.86	0.74	4.56	0.63	3.52	1.08	4.28	0.79	4.66	0.48
Description of course objectives and assignments	3.63	0.85	4.40	0.72	3.52	1.06	4.31	0.97	3.24	1.13	4.40	0.58	4.79	0.41
Requirements	3.77	0.86	4.33	0.71	3.66	0.86	4.24	0.79	3.32	1.07	4.20	0.58	4.48	0.69
Assignments	3.97	0.93	4.67	0.73	3.69	0.93	4.62	0.68	3.40	1.12	4.36	1.08	4.93	0.26
Materials	3.93	0..83	4.67	0.68	3.69	0.89	4.38	0.73	3.40	1.12	4.16	0.80	4.76	0.58
Testing procedures	3.90	0.96	4.63	0.49	3.69	0.93	4.55	0.78	3.36	1.08	4.56	0.77	4.76	0.44
Grading practices	3.80	1.06	4.73	0.58	3.66	1.08	4.76	0.51	3.36	1.11	4.72	0.54	4.90	0.31
Student Work	3.90	0.92	4.27	0.78	3.72	0.96	4.24	0.79	3.32	1.07	4.44	0.65	4.86	0.35
Average gain from pretest to posttest			0.68				0.77				1.03			
Teaching Skill														
Class meetings	4.10	0.96	4.73	0.45	3.93	0.88	4.79	0.49	3.40	1.15	4.36	0.86	4.82	0.76
Provides feedback	3.93	0.94	4.83	0.46	3.97	0.91	4.86	0.44	3.40	1.19	4.56	0.58	4.86	0.35
Instructor's command of the subject	4.07	0.94	4.97	0.18	4.07	0.96	4.93	0.26	3.72	1.06	4.64	0.49	4.86	0.35
Communication of ideas and information	3.83	0.91	4.80	0.41	3.83	0.71	4.69	0.54	3.24	1.13	4.20	0.82	4.76	0.44
Stimulation of interest in the course	3.90	0.84	4.80	0.41	3.83	0.85	4.72	0.53	3.32	1.11	4.04	0.79	4.72	0.59
Facilitation of learning	3.77	0.82	4.60	0.62	3.69	0.89	4.66	0.61	3.48	1.12	4.40	0.71	4.72	0.45
Respect and concern for students	4.00	0.88	4.87	0.35	3.93	0.96	4.86	0.58	3.56	1.00	4.60	0.76	4.86	0.44
Availability to assist students in or out of class	3.80	0.85	4.70	0.53	3.83	0.97	4.72	0.59	3.16	1.11	4.48	0.77	4.83	0.38
Average gain from pretest to posttest			0.86				0.89				1.00			

Comparisons were made on post-organization and post-teaching skills using analysis of covariance with Tukey as a follow-up procedure since all comparisons were of interest; significant group differences were not detected on the organization scores [$F(2,80) = .02$, $p > .05$], but were detected on the teaching skill scores [$F(2,80) = 5.32$, $p < .05$]. For the teaching skill scores, participants who received training on recency effect and kept a diary and those who kept a diary only had significantly different scores from the control group; there was no significant difference between the two treatment groups.

Table 2 contains the means and standard deviations for the four groups and it was interesting to note that the second control group had higher post-test scores than the other three groups since they were exposed to the course evaluation only at the end of the term; however, the assumption was made that the result was due to familiarity with the course evaluation form. We hypothesized that students who received awareness training in recency effect and kept a course diary, or kept a course diary only, would demonstrate significantly different course organization and teaching skill scores than students who did neither. Even though organization scores did not differ between the three groups, the first control group did have the greatest pre- to post-test gain, followed by the diary only group, and the training and diary group (Table 1); and the same order was found with the teaching skill scores, which may suggest that the training did sensitize students and make them more aware of recency effect as a source of rating error because their evaluations were more stable over the term. Effect sizes for teaching skill were determined for the two treatment groups and were found to be an average of .63, which also suggests that the impact of training is linked more to the behavior of the teacher rather than organizational aspects of the course.

Table 2: Means and Standard Deviations of Pre- and Post- Organization and Teaching Skill by Group

		Pre				Post			
		Organiz-ation		Teaching Skill		Organiz-ation		Teaching Skill	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training/Diary	30	30.73	6.53	31.40	6.34	35.70	3.69	38.30	2.18
Diary Only	29	29.48	6.20	31.07	6.28	35.66	4.10	38.24	2.75
Pretested Control	*25*	26.92	7.91	27.28	8.05	35.12	4.67	35.28	4.75
Non-Pretested Control	29					38.14	2.23	38.45	2.53

Qualitative Analysis

A semi-structured interview was given to 15 randomly selected participants and 6 faculty members to determine awareness of recency effect as a source of rating error and its potential impact on student evaluations of instruction (Table 3). There were 9 male and 6 female participants, who ranged in age from 20 to 28 years, and included 7 seniors, 5 juniors and 3 sophomores. All of the students had evaluated a faculty member at the university and were familiar with the course evaluation form. The key component in sensitizing to recency effect is documentation and the ability to refer to the documentation when conducting the appraisal. In this study, students were asked to update their diaries once a week at a minimum, focusing on what went well, what went poorly, and specific good or bad behaviors that were demonstrated by the instructor. To insure treatment fidelity, the researcher focused on motivation and provided periodic updates and promised to contribute $100.00 to the university's student union, and the students who participated in this study agreed to this arrangement. Documented observations made by the students, however, reflected a diverse range from one-sentence entries to several sentences concerning the instructor's performance When asked if they knew what recency effect was, two students responded that they had heard of recency effect, but only one could define the term correctly. The student who defined recency effect correctly was a

Reserve Officer Training Corps cadet who had been taught about recency effect as a rating error while in the United States Army. Once given the definition of recency effect, all 15 students believed recency effect could exist in student evaluations of faculty. Twelve of the students were able to give examples of recency effect, such as dropping of the course's lowest grade, the instructor giving a pizza party, or if by obtaining a certain average not having to take the final exam. Eleven of the students stated they had experienced recency effect, even though only one student stated he was influenced by the instructor's change of behavior.

Unsolicited comments from students included questioning the relevancy of the evaluations. All 15 students felt that there was no relevancy behind the evaluations and no matter what they wrote, it did not matter, even though two of the students commented they used past evaluations to help them select future instructors and courses. One student stated she kept a diary of all of her instructors' performance over the semester to aid her in evaluating the instructors at the end of the semester. Thirteen of the students commented that they felt they gave enough time and effort to properly evaluate their instructors, even though they felt their classmates did not. In addition to the diverse range of documentation mentioned earlier, not all diaries had a weekly entry, evidence which supports the students questioning the relevancy of the evaluations. Again, this brings into question how relevant students' think the course evaluation process is, and suggests confounding between recency effect and leniency error; still, the issue seems to be one of convincing students that their opinions do matter (Spencer & Schmelkin, 2002).

Six faculty members were interviewed regarding their knowledge of recency effect and its influence. The 6 faculty members consisted of 2 tenured instructors, 2 non-tenured instructors, and 2 adjuncts and years of teaching experience ranged from 4 to 30 years. Both tenured faculty members were aware of recency effect and how it could be used to influence an evaluation; in addition, both felt strongly that an instructor should go firmly by the syllabus, establishing objectives and requirements from the onset of the course. Three of the 4 non-tenured and adjunct professors were aware of recency effect and all 4 believed a student could be manipulated, and had thought to change their behaviors during the evaluation period to try to possibly influence their evaluation. Behaviors included dropping the lowest grade or just being more friendly or joking. One instructor did comment that he has dropped the lowest grade, even though not defined in the syllabus, in an effort to increase the students' scores to help them earn a higher grade in the course.

Unsolicited comments from the faculty members included noting the lack of maturity of students to properly evaluate an instructor's effectiveness and knowledge of course material. Five of the 6 instructors commented that the course evaluation should be used as an information-gathering tool for teaching development, but should not be used for administrative decisions; and that other means of information on faculty performance should be used to make decisions about promotion, tenure, and administrative decisions.

CONCLUSIONS

The purpose of this study was to determine whether students' were aware of recency effect and if receiving training and/or keeping a diary impacted an end-of-course evaluation. Keeping a diary and being able to reflect upon it when evaluating an instructor does seem to make a difference in a student's evaluation of an instructor as evidenced by the stability of the scores for both organization and teaching skill. Even though instruction and documentation on recency effect was provided, as well as reinforcement throughout the semester, it did not make a significant difference when compared to those who kept a diary only; which suggests that keeping the diary alone should be enough to make students aware of this source of rating error (Carrell et al.; Greenberg, 1987; Maroney & Buckley, 1992; Roberts, 1995). Overall, the mean scores for both pretest and posttest organization of course were lower than the mean scores for teaching skill, and may be explained by the students being informed and reminded to focus on the instructor and their performance. Teaching skill is an observable behavior, and the results also suggest that the impact of training is linked more to what is observable rather than structural aspects of the course. The information on the faculty member's performance is based on what the students know firsthand regarding the extent to which the faculty member actually delegates, communicates, leads, manages, and teaches (Berk, 1986); criteria regarding course organization is somewhat ambiguous and not easily observed. Organization seems to be more rigid

Table 3: Interview Responses for Recency Effect of Students and Faculty

Responses	Students	Faculty
Heard of recency effect	2	5
Able to correctly define recency effect	1	5
Believed recency effect exists in faculty evaluations	15	5
Provided a real life example	12	5
Changed behaviors prior to the evaluation to influence the evaluation (faculty only)		4
Perceived an instructor changing behavior prior to administering the evaluation to influence the evaluation (student only)	11	
Influenced by the instructor's behavior (student only)	1	

Note. Students: n = 15; Faculty: n = 6.

than teaching skill, and it could be that students perceive many aspects of course organization as pre-determined and not subject to change (Henderson, 1984; Mount & Thompson, 1987).

Whether students and faculty were aware of recency effect as a source of error that could influence end-of-course evaluations was addressed in the interview portion of the study. The semi-structured interviews with students and faculty revealed that most knew what recency effect was once it was defined. Even though most of the students had neither heard of, nor could define recency effect, once it was defined they clearly believed that such an error exists, even though they felt they could not be influenced by a change in faculty behavior. The majority of students were able to provide examples of recency effect and give real life experiences in the classroom of instructors changing their behaviors at the end of the rating period. Whether the students realized this was an attempt to influence the evaluations or not, faculty that were interviewed stated they clearly understood recency effect; 4 of the 6 faculty interviewed stated they had tried to use this phenomena to influence their evaluations at the end of the semester. Interestingly, what made the use of recency effect by faculty unique was that the four instructors identified who admitted trying to manipulate their evaluations at the end of the rating period by either behavior changes or a change to the syllabus were either non-tenured or adjunct instructors; the tenured faculty clearly stated there was no need to alter one's behavior or change the syllabus to influence an evaluation.

DISCUSSION

Students who kept a diary and used it as a source of information did demonstrate more stable course evaluation scores and appeared to be sensitized to recency effect, and supports previous findings (Carrell et al.; Greenberg, 1987; Maroney & Buckley, 1992; Roberts, 1995). It is also concluded that instructors may undergo favorable activities towards the end of a course to manipulate their evaluations if they perceive they need to, and because recency effect is an understandable error that is based on the difficulty to remember performance that may be 4 to 5 months old, one who is being rated may become more concerned about performance as the formal appraisal time approaches (Mathis & Jackson, 1994). Because of this phenomenon, individuals being rated can influence or take advantage of the recency issue by carrying out favorable acts prior to the evaluation (Mathis & Jackson).

A limitation of this study is that instructor behavior was not examined; however, the intent was to focus on the behavior of the students, consistent with the shift in emphasis of the research in this area (Chen & Hoshower, 2003; Filak & Sheldon, 2003; Spencer & Schmelkin, 2002). The results of this study indicated that the scores were not as inflated over the term when explanation and training was received, yet score direction needs to be examined. Students who are familiar with recency effect and recognize favorable acts at the end may actually tend to be more punitive, resulting in lower yet possibly more valid scores. We intend to manipulate instructor behavior in the future in order to examine if lower scores do indeed occur; however, finding an instructor who would be willing to

participate in such an experiment may prove to be quite a challenge.

Although not of interest in this study, another rating error noted, leniency, is a justification by the rater to give high marks in the evaluation based on inexperience or inability. As already stated, this rating error might have been confounded with recency effect in the evaluation since the students were adamant that there was little or no relevancy to them and that most faculty paid little attention to them (Centra, 1993; Spencer & Schmelkin, 2002); therefore, suggesting it may be decided beforehand that it is easier just to give high marks than to have to justify the evaluation to themselves or others (Carrell et al.). The unfortunate consequence of this rating error is that most of the value of systematic performance appraisal is lost (Carrell et al.; Cascio, 1982; Mathis & Jackson, 1994); this error causes scores to be inflated and results in positive skewness of the distributions. Results from this study suggest that recency effect, which appears to easily confound with other types of error, may prove difficult to examine and influence in future research.

Emphasis on behavior in rater training produces more accurate and fair evaluations, increases validity and reliability, decreases legal liabilities, and increases the overall effectiveness of the evaluation system (Carrell et al.; Cascio, 1982; Maroney & Buckley; Roberts, 1995; Turner & Clift, 1988). A key part of motivation is in understanding the relevancy of performing an activity; therefore, students need to be informed, at a minimum, of the relevancy behind student evaluations of faculty. Regardless of the source that will use the information, students need to understand what the appraisal accounts for, and if for tenure and promotion decisions, then the students need to be made aware of this. At a minimum, faculty at the beginning of the semester should inform students that they will be conducting an evaluation of the instructor's performance and should provide a copy of the evaluation form, should discuss the relevancy of the evaluation and what it is used for, and discuss keeping a diary if need be. Also, to focus more specifically on further implications and the results of recency effect, it is also recommended that future studies examine the exact behaviors exhibited by faculty that could influence an evaluation.

REFERENCES

Abbott, R.D., Wulff, D.H., Nyquist, J.D., Ropp, V.A., & Hess, C.W. (1990). Satisfaction with processes of collecting student opinions about instruction: The student perspective. *Journal of Educational Psychology*, 82(2), 201-206.

Basow, S., & Howe, K. (1987). Evaluations of college professors: Effects of professors'sex-type and sex and students' sex. *Psychology Reports, 60,* 671-678.

Berk, R. (1986). *Performance assessment: Methods and applications*. Baltimore, MD: The Johns Hopkins University Press.

Bernardin, H., & Beatty, R. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent.

Carrell, M., Elbert, N., & Hatfield, R. (1995). *Human resource management: Global strategies for managing a diverse work force* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Cascio, W. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Reston.

Cascio, W. (1989). *Managing human resources: Productivity, quality of work life, profits* (2nd ed.). New York: McGraw-Hill.

Centra, J. (1993). *Reflective faculty evaluation*. San Francisco: Jossey-Bass.

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

Centra, J., & Creech, F. (1976). *The relationship between students, teachers, and course characteristics and student ratings of teacher effectiveness*. Princeton, NJ: Educational Testing Service.

Chen, Y., & Hoshower, L.B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education, 28*(1), 71-88.

Filak, V.F., & Sheldon, K.N. (2003). Student psychological need satisfaction and college teacher-course evaluations. *Educational Psychology, 23*(3), 235-247.

French-Lazovik, G. (1982). *Practices that improve teaching evaluation*. San Francisco: Jossey-Bass.

Gall, M. D., Borg, W.R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). White Plains: Longman.

Greenberg, J. (1987). Using diaries to promote justice in performance appraisals. *Social Justice Research, 11*, 219-234.

Henderson, I. (1984). *Performance appraisal.* Reston, VA: Reston.

Maroney, B., & Buckley, R. (1992). Does research in performance appraisal influence the practice of performance appraisal? Regretfully not. *Public Personnel Management, 21*, 185-196.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707-754.

Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*(3), 253-388.

Mathis, R., & Jackson, J. (1994). *Human resource management* (7th ed.). New York: West.

McKeachie, W. (1979). Student ratings of faculty: A reprise. *Academe, 65,* 384-397.

Mohrman, A., Jr., Resnick-West, S., & Lawler, E., III. (1989). *Designing performance appraisal systems.* San Francisco: Jossey-Bass.

Mount, M. K., & Thompson, D. E. (1987). Cognitive categorization and quality of performance ratings. *Journal of Applied Psychology, 72,* 340-246.

Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology, 68,* 102-114.

Ory, J.C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching & Learning, 87,* 3-15.

Roberts, G. (1992). Linkages between performance appraisal system effectiveness and rater and ratee acceptance: Evidence from a survey of municipal personnel administrators. *Review of Public Personnel Administration, 12*(2), 19-41.

Roberts, G. (1995). Developmental performance appraisal in municipal government. *Review of Public Administration*, 15(3), 17-43.

Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education, 21,* A40.

Spencer, K.J., & Schmelkin, L.P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, *27*(5), 397-409.

Tang, S. (1999). Student evaluation of teachers: Effects of grading at college level. *Journal of Research and Development in Education, 32*(2), 83-88.

Tatro, C. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education, 28*(3), 169-174.

Turner, G., & Clift, P. (1988). *Studies in teacher appraisal.* London: The Falmer Press.

Citation

Dickey, David & Carolyn Pearson(2005). Recency Effect in College Student Course Evaluations. *Practical Assessment Research & Evaluation*, 10(6). Available online: http://pareonline.net/getvn.asp?v=10&n=6

Contact

Carolyn Pearson, Professor
University of Arkansas at Little Rock
College of Education
2801 South University Avenue
Little Rock, AR 72204

email: lcpearson@ualr.edu