

# Do Students Have Fixed Classroom Perceptions?<sup>1</sup>

**J. Ross Pruitt<sup>2</sup>**  
**Louisiana State University Agricultural Center**  
**Baton Rouge, LA 70803**



**Michael R. Dicks<sup>3</sup> and Daniel S. Tilley<sup>4</sup>**  
**Oklahoma State University**  
**Stillwater, OK 74078**

## Abstract

The importance of expectations and lack of change in student evaluation of teaching scores are key criticisms of the evaluation instrument even though prior research shows students are able to separate expectations from the final evaluation. Our research shows significant changes do occur when results are segregated by course division level. Those changes are often small, one-unit positive or negative changes from initial student perceptions leading to an average score reflecting no change. Nearly half of students change their overall instructor appraisal across the semester.

## Introduction

Students' evaluations of teaching are a common practice at institutions of higher education throughout the U.S. although the importance placed on results varies by institution. The importance placed upon the Student Evaluation of Teaching (SET) reflects the debate over whether evaluation scores accurately reflect the course experience and the persistence of expectations in determining the final evaluation. The validity of the SET is often questioned in the SET literature as a result of the belief that students form lasting opinions of instructors with only limited interaction.

Merritt argues that evaluations measure snap judgments that occur at the beginning of the semester. For teaching faculty throughout the agricultural disciplines whose SET scores are a part of the review, promotion, and tenure decision, a bad first impression could persistently plague the instructor as negative expectations would develop among current and future students. Faculty with primarily extension appointments would face a similar situation when working with producers even though a formal evaluation instrument may not be involved. If snap judgments are being recorded through SET and other evaluation measures, then the procedure is measuring perceptions of what the audience believes will occur and not what did occur.

The SET literature is full of analyses that conclude that expectations affect SET, but the question remains, "Does a change occur in the individual evaluation?" Statistically significant changes may occur when evaluations are disaggregated into upper and lower course divisions. Simply evaluating the means of course or instructor appraisal would mask shifts in the individual student scores. Overall instructor and course appraisal are reflective of specific characteristics that students perceive and changes in these underlying characteristics would alter final ratings of instructor and course appraisal. This study includes an analysis of the differences and similarities of these specific SET characteristics (for example, instructor presentation of material) between initial and final student evaluations with respect to both the instructor and the course. Our research is not designed to provide an alternative to the SET, but rather to use the SET to better understand how perceptions of student engagement and learning change over the course of the semester.

## Background

The role of students' evaluations as measures of instructor effectiveness has an extensive history of research in the United States, with nearly 2,000 published studies (Wilson, 1998). Regardless of the debate on the appropriateness of the SET as a measure of teaching quality, the SET is a tool for students to express their views on instructors and courses. Previous experiences shape the student's view of instruction and the lack of a broad educational experience, especially in lower division courses, results in lower validity of the evaluation instrument (McKeachie, 1997). As a result, comparison of the SET cannot be made across disciplines or levels (such as graduate versus undergraduate or lower division courses versus upper division courses).

Merritt (2008) states that standard SET are constructed to rely on instinctive judgments that can be formed with as little as five minutes of interaction

<sup>1</sup>The authors express their gratitude for the comments of three anonymous reviewers. All errors are those of the authors.

<sup>2</sup>Assistant Professor, Department of Agricultural Economics and Agribusiness, 101 Agricultural Administration Bldg; Phone: 225-578-2719; Fax: 225-578-2716; E-mail: rpruitt@agcenter.lsu.edu

<sup>3</sup>Professor, Department of Agricultural Economics, 314 Agricultural Hall; Phone: 405-744-6163; E-mail: michael.dicks@okstate.edu

<sup>4</sup>Professor, Department of Agricultural Economics, 422 Agricultural Hall; Phone: 405-744-6180; E-mail: daniel.tilley@okstate.edu

## Do Students

with the instructor. Widmeyer and Loy (1988) find that terms such as “warm” and “cold” to describe a guest lecturer impacts how an audience perceives and evaluates that lecturer at the conclusion of the lecture. This is consistent with persons holding to initially formed hypotheses and misinterpreting new information to support initial perceptions (Rabin and Schrag, 1999). Instructor ratings from students with *a priori* knowledge of the instructor are no more consistent than those with limited or no *a priori* knowledge (Kohlan 1973). Kohlan suggests this may reflect that students gain little new information on the ability of the instructor following initial class meetings.

Pruitt, Dicks, and Tilley (2009) find that perceptions of instructors are most impacted by “presentation of material,” for both upper and lower division courses. “Ability to explain subject matter” and a “positive attitude toward students” were also important to students in upper division courses. The “instructor's effort devoted to teaching” is also important to students in lower division courses. Perceptions of the course at the beginning of the semester were impacted most by the degree to which the individual felt the course was worthwhile for both lower and upper division courses. Other significant factors for both course divisions were perceptions about testing and evaluation procedures and if students were adequately involved in the course. These findings are consistent with Remedios and Lieberman's (2008) findings that courses perceived to be stimulating, interesting, and useful, largely determine course ratings. Remedios and Lieberman (2008) find that grades, study hours, and perceived difficulty do have a small impact on ratings.

The stability of student ratings across time is well documented (Bejar and Doyle, 1976; Costin et al., 1971; Frey, 1976; Merritt, 2008; Wetzstein et al., 1984) in the SET literature. Costin et al. (1971) note that faculty members' peer evaluations of an instructor vary across time whereas the corresponding student ratings are stable. Frey (1976) concludes that results are not “reliably different” when a subgroup of the class completes the SET at the end of a semester and another subgroup completes the SET during the first week of the following term.

Bejar and Doyle (1976) find expectations and evaluations are similar, but students are able to separate their expectations from the SET. Students in their research did not know the identity of the instructor as the pre-evaluation was administered prior to the students seeing the instructor. Wetzstein et al. (1984) compare the pre- and post-evaluations of a professor with a teaching reputation and a graduate student instructor with no teaching reputation. Using a Bayesian method, the authors conclude that the graduate student outscored the professor on the end of semester SET. Larger standard deviations are observed in the pre-evaluation than the post-evaluation which is consistent with Kohlan (1973).

## Conceptual Model

Students evaluate their instructors and courses on several different factors. Merritt (2008) suggested that students form expectations (opinions) of both courses and instructors prior to the first day of class or within the first course meetings. These expectations may be formed from input from other students, websites, professors, and/or advisors in addition to the student's own prior interaction with the instructor. These factors are in addition to the learning process students develop over their educational experiences (McKeachie, 1997).

Actual experiences in the classroom may or may not alter the student's perceptions of the course and/or instructor. No change in the student ratings indicates that information obtained from various sources regarding the instructor/course is consistent with actual experiences, that is, the expected utility from the course  $E(U_{initial})$  is equal to the actual utility received  $E(U_{final})$ . Due to the finding of Bejar and Doyle (1976) that expectations and final evaluations are consistent, but not the same, we ask if divergence occurs between expectations and actual experience?

The hypothesis is that students' experiences in the classroom and with the instructor are consistent with the expectations formed prior to taking the course and perceptions developed in the first few course meetings. More succinctly, (1)  $\Delta E(U) = E(U_{final}) - E(U_{initial}) = 0$ .

The expected utility is observed for both the overall instructor and course appraisal. The design of the SET provides specific questions related to the instructor that focus the student's mind on instructor performance prior to asking the student to appraise the overall performance of the instructor. A similar pattern is followed for questions related to the course.

## Data

Students in twenty-two courses in the College of Agricultural Sciences and Natural Resources (CASNR) participated in this research. Classes with students participating were from the departments of animal science, agricultural economics, agricultural communication, education, and leadership, plant and soil science, horticulture, and natural resource ecology and management. Of the twenty-two courses, one course was being offered for the first time by a full professor and one course was being taught by a graduate instructor for the first time. Two freshmen, four sophomore, nine junior, and seven senior level courses were used. Seventeen instructors participated with nine of those being full professors. Of the remaining instructors, two were associate professors, five were assistant professors, and one was a graduate student instructor.

Evaluations were completed within the first two weeks of the fall 2007 semester with evaluation time being determined by the instructor to allow for the least amount of intrusion to the instructor. As a

reviewer noted, this time span may impact initial perceptions, but an instructor's desires to have (freshmen) course rosters and logistics settled before the first instructional period led to a wider variation in the date of the initial SET than desired. The final round of evaluations was conducted from November 15th through December 7th. Students, on average, completed the questionnaire in fifteen minutes at both points in the semester.

An informational cover sheet was included that listed the title of the research, a student's rights as a research volunteer, instructions on how to determine their individual identification code, and a statement that the research would occur twice in the semester. The individual identification code is used to match responses at the beginning with those at the end of the semester while maintaining anonymity and confidentiality. The code number is a five digit alphanumeric code based on information known only to the student. The first digit is the first letter of the high school where the student graduated. Digits two and three are the student's birth month (January is 01, February as 02, and so forth) with the final two digits being the last two digits of the student identification number. The informational cover sheet also indicated the confidentiality of all responses including the fact that instructors would not see the results until after grades had been submitted.

A total of 867 evaluations were collected at the beginning of the semester and 897 evaluations were collected at the end of the semester. A total of 423 evaluations were successfully matched by identification code number from earlier in the semester. Sixty-two percent of respondents whose SET were matched

were females compared to 47% of persons who completed the initial SET. Students majoring in a field contained in CASNR accounted for approximately 95% of responses. Eighty percent of respondents reported the course was required and students with junior standing were the mean class.

## Results

Aggregation of all matched responses for mean overall instructor appraisal shows no statistical difference between the beginning and end of the semester. However, statistically significant changes do occur when responses are segregated by course division. Comparison of means by a pooled means test for upper and lower division courses are included in Tables 2 and 3, respectively. Statistically significant decreases in the mean overall appraisal of instructors occur in upper division courses while statistically significant increases in lower division courses occur. For students in upper division courses, two of the factors related to the instructor show a statistically significant increase (*Knowledge and Attitude*) and two show a significant decrease (*Presentation and Explain*) compared to four statistically significant increases for students in lower division courses. This suggests that students are perceptive of difference in specific instructor and course factor characteristics. Changes in factor characteristics in turn alter the way students perceive overall instructor and course appraisal. For example, the decline in overall instructor appraisal among students in upper division courses may be a result of the student's perception of the decline in ability to explain and present material.

Presentation of material is the leading factor that determines differences in perceptions of initial instructor appraisal (Pruitt, Dicks, and Tilley, 2009) and the decline of this characteristic likely impacted the overall instructor appraisal score.

Statistically significant differences in means were found for all course characteristics in both upper and lower division courses. This is largely due to students having the option of rating a course characteristic as undecided or not applicable. Undecided or not applicable responses were grouped together at the beginning of the semester and treated as being in the middle of the rating scale. The "not applicable" option was included on the final SET, but responses that marked this option were treated as non-responses at the end of the semester due to the presence of the undecided option.

Observed standard deviations increase with many of the collected factors and factor characteristics which stands in contrast to the findings of Wetzstein et al. (1984). The majority of instructor and course characteristics largely see increases in standard deviations across the semester and may be par-

Table 1. Questions on University Student Evaluation of Teaching Form

Variable Name	Question
Instructor Variables	
<i>Preparation<sup>a</sup></i>	Preparation and effort
<i>Teaching Effort<sup>a</sup></i>	Effort devoted to teaching
<i>Presentation<sup>a</sup></i>	Presentation of material
<i>Knowledge<sup>a</sup></i>	Knowledge of subject
<i>Explain<sup>a</sup></i>	Ability to explain subject matter
<i>Attitude<sup>a</sup></i>	Positive attitude toward students
<i>Instructor Overall<sup>f</sup></i>	Overall INSTRUCTOR appraisal
Course Variables	
<i>Workload<sup>b</sup></i>	The workload is appropriate for the hours of credit
<i>Assignments<sup>b</sup></i>	Assignments are relevant and useful
<i>Tests<sup>b</sup></i>	Testing and evaluation procedures are good
<i>Involve<sup>b</sup></i>	Students are adequately involved
<i>Worthwhile<sup>b</sup></i>	This course is worthwhile to me
<i>Course Overall<sup>b</sup></i>	Overall, this is a GOOD course

<sup>a</sup> Options were Very High, High, Average, Low, or Very Low.

<sup>b</sup> Options were Definitely Yes, Yes, Undecided, No, Definitely No, or Not Applicable.



**Table 2. Comparison of Means for All Collected Questionnaire Types for Upper Division Courses**

Variable	Matched initial			Matched final	
	Range	Evaluations		Evaluations	
		Mean	Std Dev	Mean	Std Dev
<b>Instructor Characteristics<sup>z</sup></b>					
<i>Preparation</i>	0-4	3.41	0.71	3.51	0.74
<i>Teaching Effort</i>	0-4	3.49	0.65	3.57	0.70
<i>Presentation</i>	0-4	3.52	0.62	3.26*	0.95
<i>Knowledge</i>	0-4	3.26	0.78	3.60*	0.73
<i>Explain</i>	0-4	3.59	0.58	3.38*	0.92
<i>Attitude</i>	0-4	3.41	0.71	3.65*	0.66
<i>Instructor Overall</i>	0-4	3.65	0.55	3.54*	0.74
<b>Course Characteristics<sup>y</sup></b>					
<i>Workload</i>	1-5	3.83	0.77	4.27*	0.78
<i>Assignments</i>	1-5	3.83	0.78	4.29*	0.82
<i>Tests</i>	1-5	3.54	0.73	4.20*	0.93
<i>Involve</i>	1-5	4.00	0.72	4.40*	0.71
<i>Worthwhile</i>	1-5	4.00	0.81	4.25*	0.95
<i>Course Overall</i>	1-5	3.90	0.79	4.34*	0.86

<sup>z</sup> Very low is 0, Low is 1, Average is 2, High is 3, and Very High is 4.  
<sup>y</sup> Definitely No is 1, No is 2, Undecided/Not Applicable is 3, Yes is 4, and Definitely Yes is 5.  
 \*Mean is significantly different from corresponding mean in matched initial evaluation column at the 5% level

**Table 3. Comparison of Means for All Collected Questionnaire Types for Lower Division Courses**

Variable	Matched initial			Matched final	
	Range	Evaluations		Evaluations	
		Mean	Std Dev	Mean	Std Dev
<b>Instructor Characteristics<sup>z</sup></b>					
<i>Preparation</i>	0-4	3.21	0.75	3.56*	0.60
<i>Teaching Effort</i>	0-4	3.34	0.74	3.65*	0.55
<i>Presentation</i>	0-4	3.23	0.74	3.31	0.77
<i>Knowledge</i>	0-4	3.05	0.80	3.75*	0.49
<i>Explain</i>	0-4	3.50	0.73	3.47	0.70
<i>Attitude</i>	0-4	3.21	0.750	3.64*	0.64
<i>Instructor Overall</i>	0-4	3.32	0.83	3.59*	0.61
<b>Course Characteristics<sup>y</sup></b>					
<i>Workload</i>	1-5	3.41	0.62	4.24*	0.67
<i>Assignments</i>	1-5	3.40	0.62	4.24*	0.72
<i>Tests</i>	1-5	3.22	0.53	4.13*	0.86
<i>Involve</i>	1-5	3.48	0.67	4.21*	0.67
<i>Worthwhile</i>	1-5	3.59	0.69	4.14*	0.87
<i>Course Overall</i>	1-5	3.56	0.67	4.28*	0.81

<sup>z</sup> Very low is 0, Low is 1, Average is 2, High is 3, and Very High is 4.  
<sup>y</sup> Definitely No is 1, No is 2, Undecided/Not Applicable is 3, Yes is 4, and Definitely Yes is 5.  
 \* Mean is significantly different from corresponding mean in matched initial evaluation column at the 5% level

tially the result of the period of time in which final SET were collected. Increased standard deviations reflect the changes in scores of instructor and course appraisal as well as factor characteristics by the individual student. Although additional instructional periods and graded assignments could impact the results seen for course characteristics, the ability of the instructor to influence students' ability to learn and provide an experience different from the expectation does occur. Kohlan (1973) states that little new information about the instructor is gained after the initial course meetings, but this lack of new information does not mean that students are not interpreting the information in a different manner.

Cross tabulations are calculated for course and instructor variables and shown in Tables 4 and 5 (initial ratings are in the rows with final ratings in the columns). These tables show the distribution of changes by direction and magnitude of the change which illustrate why standard deviations increase while the means of collected variables show sometimes statistically insignificant changes. Note that only a small percentage (less than 5%) of the students use below average evaluations and almost all of the below average ratings were observed at the end of the semester. Of the 234 evaluations collected in upper division courses, 50 students decrease their overall instructor appraisal rating, 32 increase their instructor appraisal, and 152 do not change. In lower division courses, 29 students decrease their overall instructor appraisal, 62 increase instructor appraisal, and 98 do not change among 189 observations. For overall course appraisal in upper division courses, 30 students decrease their ratings, 102 students increase their opinion, and 100 exhibit no change of overall course appraisals. Overall course appraisal in lower division courses had 14 students decrease their ratings, 119 increased their ratings, and 54 students exhibited no change.

The majority of changes shown in Tables 4 through 7 are one unit changes in either direction although larger changes are observed for course appraisal due to the presence of the undecided/not applicable option at the beginning of the semester. When graphed, the changes that do occur approximate a normal distribution which results in the charge that SET are constant throughout the semester. The rating may not change, but as Bejar and Doyle (1976) show,

**Table 4. Cross Tabulation of Overall Instructor Appraisal across the Semester in Upper Division Courses**

		End of Semester					Total
		Very Low	Low	Average	High	Very High	
Beginning of Semester	Very Low	0	0	0	0	0	0
	Low	0	0	0	0	0	0
	Average	0	1	3	1	4	9
	High	0	1	10	25	27	63
	Very High	1	1	7	29	124	162
	Total	1	3	20	55	155	234

Note: The question stated "Overall INSTRUCTOR appraisal"

**Table 5. Cross Tabulation of Overall Course Appraisal across the Semester in Upper Division Courses**

		End of Semester					Total
		Definitely No	No	Undecided	Yes	Definitely Yes	
Beginning of Semester	Definitely No	0	0	0	0	0	0
	No	0	0	0	0	0	0
	Undecided/Not Applicable	2	7	12	31	34	86
	Yes	0	1	8	38	37	84
	Definitely Yes	0	1	0	11	50	62
	Total	2	9	20	80	121	232

Note: The question stated "Overall this is a GOOD course"

**Table 6. Cross Tabulation of Overall Instructor Appraisal across the Semester in Lower Division Courses**

		End of Semester					Total
		Very Low	Low	Average	High	Very High	
Beginning of Semester	Very Low	0	0	0	0	2	2
	Low	0	0	0	3	0	3
	Average	0	1	0	6	16	23
	High	0	1	0	30	35	66
	Very High	0	1	3	23	68	95
	Total	0	3	3	62	121	189

Note: The question stated "Overall INSTRUCTOR appraisal"

**Table 7. Cross Tabulation of Overall Course Appraisal across the Semester in Lower Division Courses**

		End of Semester					Total
		Definitely No	No	Undecided	Yes	Definitely Yes	
Beginning of Semester	Definitely No	0	0	0	0	0	0
	No	0	1	0	0	0	1
	Undecided/Not Applicable	2	2	5	50	39	98
	Yes	0	2	2	37	30	71
	Definitely Yes	0	0	1	5	11	17
	Total	2	5	8	92	80	187

Note: The question stated "Overall this is a GOOD course"

this does not mean that expectations are equal to the final evaluation.

## Conclusions

Expectations are important in obtaining student engagement and determining the final student evaluation of teaching, but those expectations are not always equal to actual experiences of students as measured by the SET. As found in Bejar and Doyle (1976), we find significant differences between expectations measured early in the semester and later in the semester. These changes become apparent when viewed by course division level as opposed to aggregation of all matched responses. Final evaluation scores do reflect the expectations and perceptions a student begins the semester with, but ability of the instructor and course content does change the student's thinking as reflected on the end of semester SET. Students in upper division courses tend to decrease their evaluations of the instructor while students in lower division courses tend to increase their evaluations of the instructor.

Analysis of the means and standard deviations collected for course and instructor appraisal as well as their related factor characteristics, masks the large number of students who change their responses over the course of the semester. The numbers of students who change their instructor and course appraisal scores exceed those that do not, indicating a difference between expectations and actual performance of course and instructor. This suggests that instructors do have influence on students' ability to learn and leads to students having an actual experience that is different from their previously held expectation.

Changes in factor characteristic scores show significant changes across the semester indicating that students are perceptive of differences and will not hold to their original expectation of a specific factor characteristic. Changes in perceptions about a specific factor characteristic can certainly lead to changes in overall instructor or course appraisal. Statistically significant decreases in the mean score of presentation of material by an instructor in upper division courses as the semester progresses suggests that as material becomes more difficult through a semester, instructors should alter the manner of their presentation to find new ways to connect with students. Students in those upper division courses do not see an increase in the effort devoted to teaching as the semester progresses, and this may be nega-

## Do Students

tively impacting the presentation of material score. Students in lower division courses do see a difference in effort devoted to teaching across the semester and although the presentation of material does not change, overall instructor appraisal does see an increase.

When a student enters a classroom, the previous educational experiences also enter and shape the perceptions of that student. Students with poor expectations for the instructor and the course may be less engaged and perform at a lower level than students with higher expectations. Other factors shape expectations, especially of items the students would not have experience with such as course exams and assignments. The lack of experience with course exams and assignments does not prevent students in upper division courses from expressing a perception or expectation on these items. In our data, the mean rating is closer to 4 (agreeing with appropriate evaluation instruments) as opposed to being neutral or undecided (a rating of 3). In lower division courses, the initial ratings were more neutral or undecided. How these expectations are being formed should be a topic of future work since Pruitt et al. (2009) found that recommendations from friends, professors, and websites often have little explanatory power (specifically on presentation of material, effort devoted to teaching, and the worth of the course).

Each instructor needs to carefully evaluate, throughout the semester, those changes in perceptions as this research documents, students can be negatively or positively impacted by what occurs in the classroom. Some of these perceptions can be managed while still expectations, that is, prior to any contact by the student with the instructor and course. The SET literature possesses many ways to alter these perceptions (Merritt, 2008; Wilson, 1998) although caution should be used as these strategies may not improve student learning.

Although this research does shed light on changes that do occur from initial perceptions and expectations of students on instructor and course characteristics, the underlying factors that lead to these changes are not understood and should be a course of future research. Remedios and Lieberman (2008) do indicate that stimulating courses that are interesting and useful largely determine course ratings and would likely impact the appraisal of the instructor.

## Literature Cited

- Bejar, I.I., and K.O. Doyle. 1976. The effect of prior expectations on the structure of student ratings of instruction. *Jour. of Educational Measurement* 13(2):151-54.
- Costin, F., W.T. Greenough, and R.J. Menges. 1971. Student ratings of college teaching: Reliability, validity, and usefulness. *Rev. of Educational Research* 41(5):511-35.
- Frey, P.W. 1976. Validity of student instructional ratings: Does timing matter? *Jour. of Higher Education* 47(3):327-36.
- Kohlan, R.G. 1973. A comparison of faculty evaluations early and late in the course. *Jour. of Higher Education* 44(8):587-595.
- McKeachie, W.J. 1997. Student ratings: The validity of use. *American Psychologist* 52(11):1218-1225.
- Merritt, D.J. 2008. Bias, the brain, and student evaluations of teaching. *St. John's Law Rev.* 82(4):235-87.
- Pruitt, J.R., M.R. Dicks, and D.S. Tilley. 2009. Determinants of students' first impressions of instructors and Courses. *NACTA Jour.* 53(3):24-36.
- Rabin, M. and J.L. Schrag. 1999. First impressions matter: A model of confirmatory bias. *Quarterly Jour. of Economics* 114(1):37-82.
- Remedios, R. and D.A. Lieberman. 2008. I liked your course because you taught me well: The influence of grades, workload, expectations, and goals on students' evaluations of teaching. *British Educational Research Jour.* 34(1):91-115.
- Wetzstein, M.E., J.M. Broder, and G. Wilson. 1984. Bayesian inference and student evaluations of teachers and courses. *Jour. of Economic Education* 15(4):40-45.
- Widmeyer, W.N. and J.W. Loy. 1988. When you're hot, you're hot! Warm-cold effects in first impressions of persons and teaching effectiveness. *Jour. of Educational Psychology* 80(1):118-21.
- Wilson, R. 1998. New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education* January: A-12-14.