

Surprises learned from course evaluations

Wei-Kian Chen
Ohio Northern University

Won-Sun Chen
SEGi University College, Malaysia

ABSTRACT

Course evaluations are widely integrated into faculty and course assessments in higher education. Evaluations provide student-evaluated quality information and are used for tenure, promotion and salary adjustment decisions. In this study, a simple instrument for student instructor rating was used to collect evaluation data from January through December 2009. The evaluations were voluntarily submitted by 758 students. A statistical control chart was used to monitor the longitudinal performance of instructors and provide feedback to the administration regarding whether or not short-term or long-term attention and guidance are needed. Instructor's performance, as affected by both class size and the level of difficulty of the courses, was evaluated through various regression techniques. The results of this study indicated that multiple factors, principally increased course-load levels and motivational variations resulting from delivering the same course several times in a year, can lower the performance of instructors over time. Also, different types of pedagogies may also influence the outcome of the evaluations. On the other hand, larger class size does not necessary reduce the instructor's performance.

Keywords: Course evaluations, students rating of instructors, class size, statistical control chart, evaluation instrument.

INTRODUCTION

The significant increase in the number of higher education institutions over the past few decades has created an environment of active competition among education providers. Students have many options when they are identifying and selecting a quality program from a reputable college or university. Therefore, it is important to ensure that the quality of programs are maintained and accredited by the relevant qualification agencies or associations. One way of assessing the quality of the program is via student ratings of instructors (SRIs). Students are asked to evaluate their instructor on how he/she “perform” throughout the course (Goldstein & Benassi, 2006). The SRIs can also have an indirect influence on the tenure promotion and salary adjustment for instructors (McKeachie, 1979; Theall & Franklin, 1990; Yao, Weissinger & Grady, 2003).

Although there are various types of valid instruments developed for SRIs, different institutions apply different format to address the teaching and instructor. It is difficult to deny the fact that evaluating student satisfaction of teaching remains controversial and problematic (Richardson, 2005; Wiser-Jensen, Stensaker & Groggaard, 2003; Baxter Magolda, 1992). Numerous factors, such as cultural background and gender of the student, year of which the student is into his/her study, types of courses, class size, when the courses are offered, are found to be associated with the course evaluations (Davis, Hirschberg, Lye, Johnston & McDonald, 2007).

Generally, no in-depth analyses are performed on the collected evaluation data. Therefore, most of the time, the instructors only receive a brief summary report on their performance. This is because evaluation data are not easy to analyze due to its ordinal nature and usually little guidance are received in analyzing and interpreting the analyses results. In additions, the degree to which the administration incorporates this information into an overall performance appraisal of staff is often vague (Millman & Darling-Hammond, 1990; Scriven, 1995; Theall, Franklin & Ludlow, 1990; Wachtel, 1998). On the other hand, one should be cautious in reading the summary report generated from evaluation data because if students know instructors are giving them easy grades, then, they will normally give excellent evaluations in return. In some occasions, students maybe too afraid to rate their instructor according to the truth because they are worried this might jeopardize their grades. Students’ narratives, if provided, are another source of performance feedbacks for instructors.

In terms of tenure promotion and salary adjustment, the summary report from evaluation data can be used as part of the supporting documents. These reports will act as quantitative evidence on whether or not the relevant instructor deserves to be promoted or for salary adjustment.

This retrospective study aims to employ a statistical control chart approach in analyzing the evaluation data. Techniques like this not only provide good quantitative evidence to the head of department on how his/her staff perform throughout the year, but also provide a good indication to the head of department on when faculty should be advised and monitored. It is also interesting to observe the various longitudinal trends of the instructor’s performance versus the class size as well as the level of difficulty of the course, for example, freshman, sophomore, junior or senior level of courses.

INSTRUMENT

This is a generic instrument only meant for course evaluations. It comprises 11 simple questions which are designed to evaluate the instructor's ability to explain, fairness in grading, degree of preparation for classes, ability to stimulate interest, enthusiasm for the course, helpfulness and availability outside classroom, receptiveness to questions asked in class, effectiveness as an instructor, quality of teaching, and ability to present material effectively. Unfortunately, no information about the students' background are collected. In this instrument, students ought to spend approximately 15 minutes to rate their instructor on a scale of 1 (far below average) to 5 (well above average). The total score, which ranges from 11 to 55, will then be calculated as the summation of the rating for each question. In general, a higher total score indicates the instructor is well received by students. A remark column is provided at the end of the instrument for students to document their comments or suggestions.

DATA

American Degree Program students from freshman to senior level in psychology, communication, English, humanity and social science related courses were selected to participate in this study. This is a franchised program where students will spend 4 years of their study in an accredited college in Malaysia, then, students will receive their Bachelor of Science degree from a reputable American university upon completion of their study. The courses mentioned earlier were selected because they were offered in this program during the period of data collection and only complete evaluation data were included. The data were collected from a total of 35 course evaluations from 5 different terms offered between January and December 2009. These 35 courses were taught by 5 full-time instructors in the program. All these instructors possess a master qualification in their relevant field of study, majority of them have less than 2 years of teaching experience, only one instructor has more than 2 years of teaching experience. Overall, the data contained the evaluations voluntarily submitted by 758 students.

ANALYSIS

A statistical control chart known as the mean or alternatively the x-bar control chart along with its lower control limit (LCL) and upper control limit (UCL) were used to evaluate the collected data. This technique will not only provide a detail control chart which is used to monitor the instructors' performance over time, but also indicate when extra guidance and support ought to be provided. This is important especially for new instructor who may need thorough guidance and close monitoring from time to time. Besides, various types of regression techniques were applied to investigate the relationship between evaluation score versus both class size and level of difficulty for courses offered.

RESULTS

Table 1 summarized the level of difficulty for courses along with the class sizes that were included in this study. Majority of the courses offered were at freshman level with class sizes ranged from 14 to 57 students. Generally, class size decreased as the level of difficulty increased.

Figure 1 presented the evaluation results for 5 instructors in the program throughout the year of 2009. It was interesting to note the excellent improvement in lecturing for instructor 1 over time. This particular instructor started with course evaluations worse than the segregated average, but tremendous improvements were observed at the end of year 2009. On the other hand, the rest of the instructors showed a slight drop in their in-class performances, but a significant drop was observed for instructor 4, who has more than two years of teaching experience. Overall, majority of the instructors' performances were consistently above the segregated average except for instructors 1, 3 and 4.

The influence of class size on course evaluations was showed in Figure 2. It was interesting to note the in-class performance of instructors generally dropped slightly with increasing class sizes. This was the case for all instructors except for instructor 3.

A polynomial model mentioned below was developed to investigate the influence of class size on course evaluations. Figure 3 clearly showed the data were fitted using a U-shape model. This model presented the finding that course evaluations decreased slightly when class size increased to approximately 30 students, then, class size of 25 to 35 students produced the lowest evaluations, and the evaluations peaked again at class size of 30 students and beyond.

$$\text{Evaluation score} = 0.0111 \times (\text{class size})^2 - 0.6415 \times (\text{class size}) + 50.461$$

Figure 4 presented course evaluations by level of difficulty for courses offered. Again, a polynomial model was used to fit the data. Generally, this figure indicated higher level of courses did not necessary produce lower course evaluations. On average, the evaluation scores for junior level courses were slightly higher compared to freshman level courses. Below you will find the details of the model that was used to fit the data.

$$\text{Evaluation score} = 0.528 \times (\text{level of difficulty for classes offered})^2 - 2.0437 \times (\text{level of difficulty for classes offered}) + 45.655$$

According to Figure 5, it was surprising to observe the course evaluations for freshman level courses did not influence by the increasing class size. In fact, some freshman level courses produced better course evaluations with larger size. As for sophomore level courses, the course evaluations dropped when class size was larger than 35 students. Good course evaluations for junior level classes were observed when class size was less than 10 students. Again, a polynomial model mentioned below was used to fit data for each level of courses offered.

For Freshman classes:

$$\text{Evaluation score} = 0.0055 \times (\text{class size})^2 - 0.2691 \times (\text{class size}) + 46.704$$

For Sophomore classes:

$$\text{Evaluation score} = -0.113 \times (\text{class size})^2 + 7.5031 \times (\text{class size}) - 82.246$$

For Junior classes:

$$\text{Evaluation score} = -0.1544 \times (\text{class size})^2 + 2.1636 \times (\text{class size}) + 42.93$$

DISCUSSION

Since all instructors do not possess a broad experience in teaching, it is possible for their performance to drop slightly over time. Few factors, including increase of course loads and motivation, could contribute significantly to this finding.

Besides, the instructor's style of teaching is also an important factor to consider. A conventional teacher-centered pedagogy, where students are expected to blindly accept the information that the instructor provided in class without given a chance to ask questions (Stofflett, 1998), is no longer found to be effective (Lord, Travis, Magill & King, 2009). Instead, instructors should start exploring the use of student-centered pedagogy, where students' learning is the focus of the class and students are also encouraged to interact regularly with their instructors as well as to hold small group discussions among them (Yager, 1991). The use of these strategies can have a direct effect not only on students' learning and course evaluations, but also the tenure and promotion decisions (Hara, 2009).

Although class size is found to have only a slight impact on the course evaluations, this finding is not unusual because the same finding has also been observed in a past study (Marsh, Overall, & Kesler, 1979). This finding has indirectly challenged the widely held speculation that effectiveness of instruction necessary suffers in classes with larger size. Polynomial model as shown in Figure 3 has also supported the finding that performance of an instructor can be maintained or even better for courses with larger class sizes. In this study, freshman level courses are generally with larger class sizes. This could be due to the fact that freshman courses are easier and students tend to enjoy the role plays and activities that are incorporated into the courses. On the other hand, an opposite outcome is observed for higher level courses. This could due to the fact that these higher level courses are tougher and students prefer to have a smaller learning environment.

This study is limited with only one year of complete course evaluations data because no complete data are available prior to year 2009 for this program. Therefore, further studies to include more data are suggested to support the findings of this study. Since the chosen program is relatively young in terms of offering courses, therefore, it would be beneficial to include more senior level courses into future studies to allow for a regression model development.

Table 1. Types of courses included.

Level of difficulty	Types of courses	Number of courses offered	Class size	
			Minimum	Maximum
Freshman	Humanity and Social Science	7	14	57
	English	5		
	Psychology	3		
	Communication	4		
	<i>Total</i>	<i>19</i>		
Sophomore	Communication	1	29	41
	Humanity	2		
	<i>Total</i>	<i>3</i>		
Junior	Psychology	5	2	18
	Communication	1		
	Humanity	5		
	<i>Total</i>	<i>11</i>		
Senior	Communication	1	4	
	<i>Total</i>	<i>1</i>		

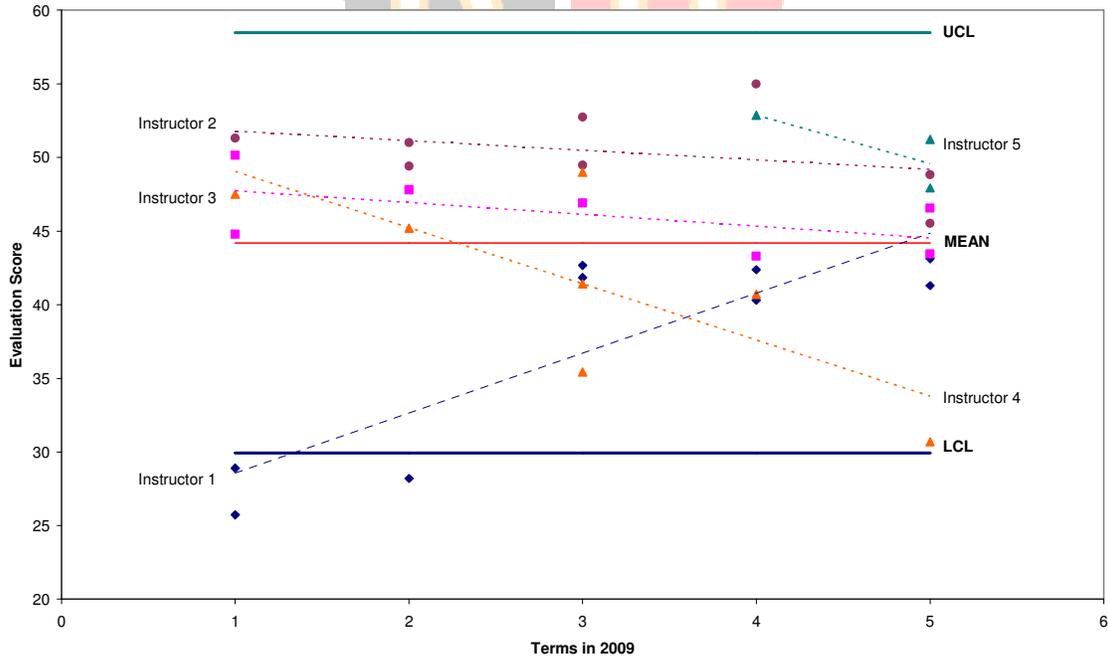


Figure 1. Evaluation scores for each instructor by term in year 2009

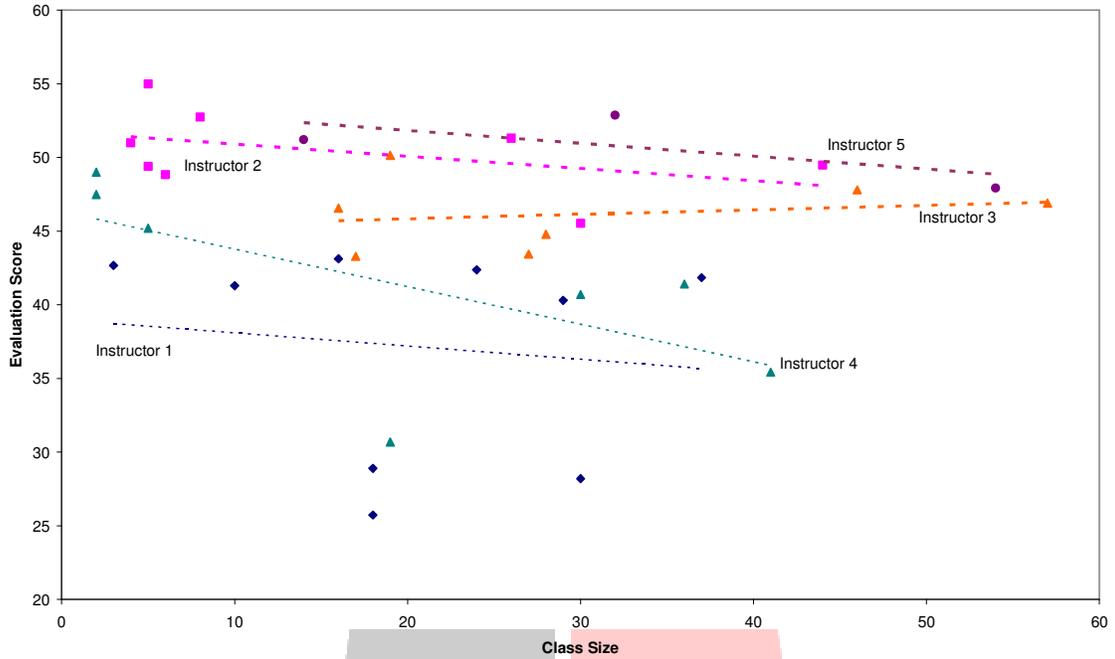


Figure 2. Evaluation scores for each instructor by class size in year 2009

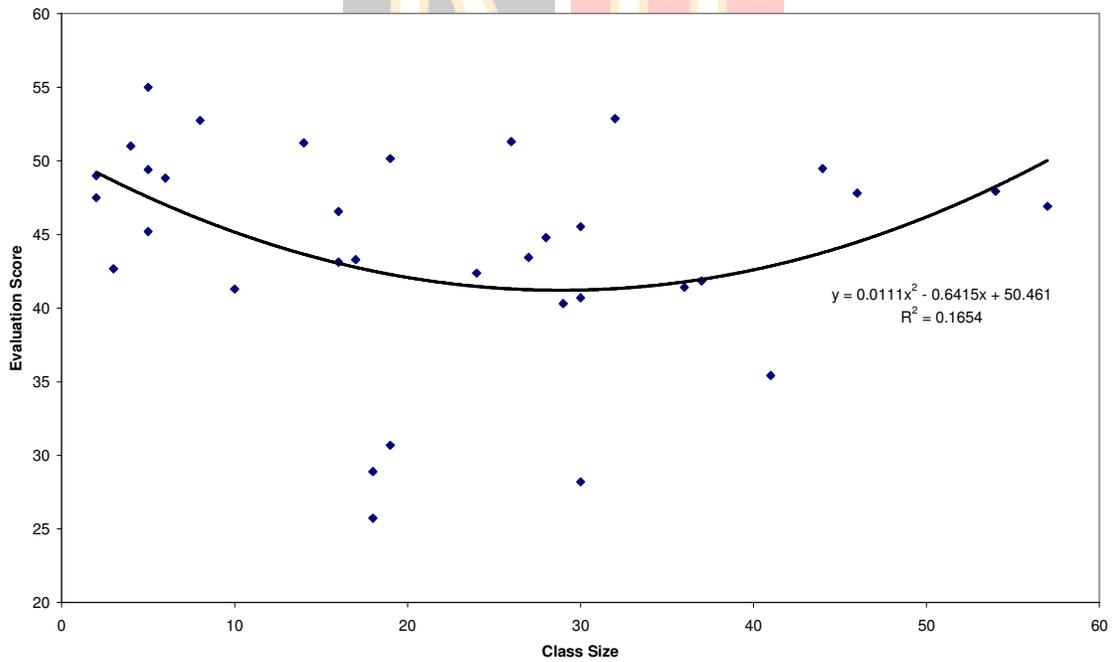


Figure 3. Modeling of evaluation scores by class size in year 2009

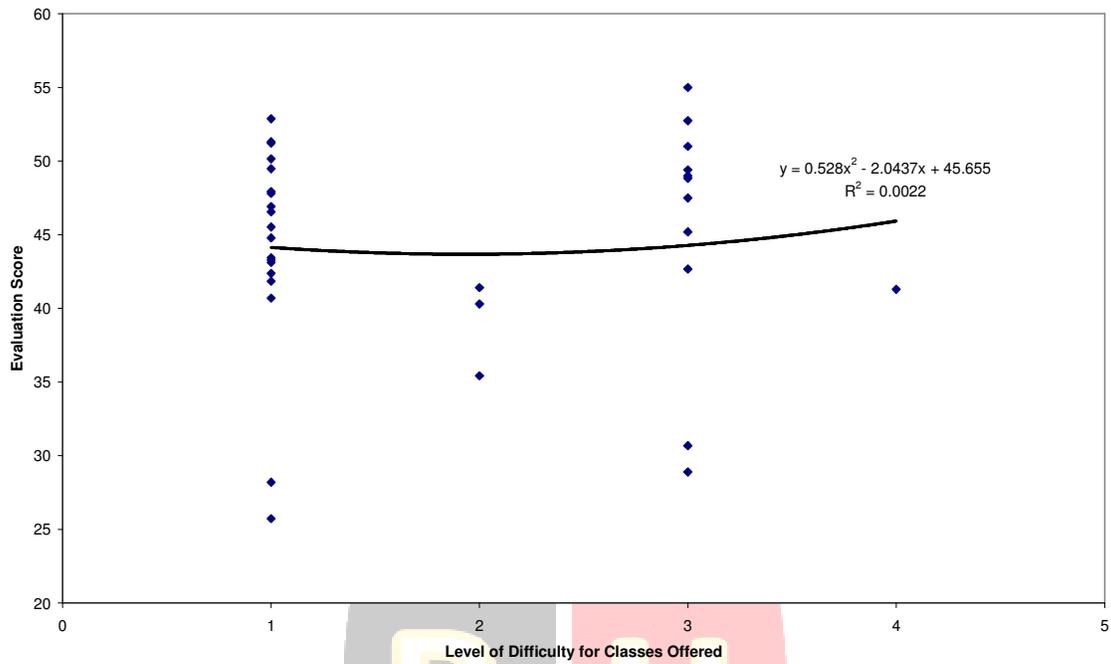


Figure 4. Modeling of evaluation scores by term in year 2009

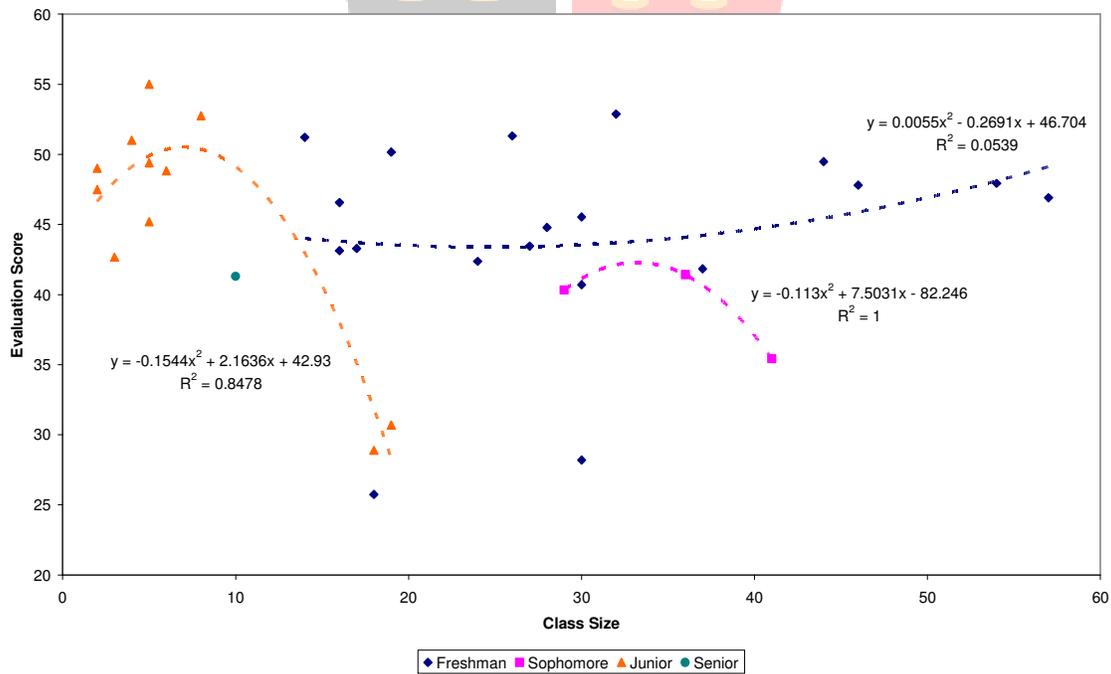


Figure 5. Modeling of evaluation scores by class size and level of difficulty for courses offered in year 2009

REFERENCES

- Baxter Magolda, M.B. (1992). *Knowing and reasoning in college: gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.
- Davis, M., Hirschberg, J., Lye, L., Johnston, C. & McDonald, I. (2007). Systematic influences on teaching evaluations: the case for caution. *Australian Economic Papers*, march, 18-38.
- Goldstein, G.S. & Benassi, V.A. (2006). Students' and instructors' beliefs about excellent lecturers and discussion leaders. *Research in Higher Education*, 47(6), 685-707.
- Hara, B. (2009). Teacher-centered vs. student-centered pedagogy. Retrieved on September 18, 2009 from <http://www.profhacker.com/2009/09/18/teaching-centered-vs-student-centered-pedagogy/>.
- Lord, T., Travis, H., Magill, B. & King, L. (2002, April). Comparing student-centered and teacher-centered instruction in college biology labs. Paper presented at Pathways to Change: An International Conference on Transforming Math and Science Education in the K16 continuum. Arlington VA: Crystal City.
- Marsh, H.W., Overall, J.U. & Kesler, S.P. (1979). Class size, students' evaluations, and instructional effectiveness. *American Educational Research Journal*, 16(1), 57-69.
- Mckeachie, W.J. (1979). Student ratings of faculty; A reprise. *Academic Bulletin of the AAUP*, 65(6), 384-397.
- Millman, J. & Darling-Hammond, L. (1990) (Editors). *The new handbook of teacher evaluation*. SAGE: Newbury Park.
- Richardson, J.T.E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment and Evaluation in Higher Education*, 30(4), 387-415.
- Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. *Practical Assessment, Research & Evaluation*, 4(7).
- Stofflett, Rene T. (1998). Putting Constructivist Teaching into Practice in Undergraduate Introductory Science. *Electronic Journal of Science Education*, Vol. 3, No. 2. Retrieved December 31, 2001 from <http://unr.edu/homepage/jcannon/ejse/stofflett.html>.
- Theall, M. & Franklin, J. (1990) (Editors). *Student Ratings of Instruction: Issue for Improving Practice*. San Francisco, CA: Jossey-Bass.
- Wachtel, H.K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23, 191-210.
- Wiers-Jensen, J., Stensaker, B. & Groggaard, J.B. (2003). Student satisfaction: towards an empirical deconstruction of the concept. *Quality in Higher Education*, 8, 183-195.
- Yager, Robert E. (1991). The Constructivist Learning Model. *The Science Teacher*. Vol. 58, No. 6:53-57.
- Yao, Y., Weissinger, E. & Grady, M. (2003). Faculty use of student evaluation feedback. *Practical Assessment Research & Evaluation*, 8(21).