



University of
Massachusetts
Amherst

Assessing Teaching Effectiveness in a Basic Foods Laboratory Setting: Phase Four - Pilot Testing the Instrument

Item Type	pap_con;event
Authors	Weber, Melvin R;Chandler, James A.;Finley, Dori A
Download date	2025-02-17 08:24:23
Link to Item	https://hdl.handle.net/20.500.14394/42613

ASSESSING TEACHING EFFECTIVENESS IN A BASIC FOODS LABORATORY SETTING: PHASE FOUR – PILOT TESTING THE INSTRUMENT

Melvin R. Weber
East Carolina University
Greenville, NC USA

James A. Chandler
East Carolina University
Greenville, NC USA

and

Dori A. Finley
East Carolina University
Greenville, NC USA

ABSTRACT

This is the fourth phase of a research project to develop a student evaluation tool (SET) to evaluate teaching effectiveness in a basic foods laboratory taught at senior institutions. The goal of phase four is to test the instrument among students enrolled in basic foods laboratories. Four hundred seventy five students at eight U.S. universities who were enrolled in a basic food laboratory completed the SET and a demographic questionnaire. Reliability using Cronbach's coefficient alpha was .962 and factor analysis yielded 3 factors comparable to those identified by the Delphi Panel. Results from the first part of pilot testing indicate a reliable instrument preferred by students enrolled in basic food laboratory classes.

INTRODUCTION

To ensure continuous improvement, providing and receiving performance feedback is a key concept in any job. Within the academy, educators are routinely evaluated by students (and their peers) using some form of evaluation tool. An evaluation tool serves many functions: (1) providing feedback, (2) identifying training and development needs, (3) making decisions linking rewards and discipline to performance, (4) evaluating policies, programs, or training, and (5) validating the selection process (Woods, 2002). Within the academic community, all five of these functions are applicable to evaluating classroom teaching.

Most educators (including the authors of this research) would contend that different evaluation tools should be implemented in the various types of classroom environments. A traditional, lecture style class is different than a laboratory setting. It follows, then, that these two disparate types of learning environments should not use the same teaching effectiveness evaluation tool. Frequently the same evaluation form is used, with the exception that the word "classroom" is replaced with "laboratory." The most critical question is: does this simple change allow for a truly accurate and fair evaluation of instruction in the laboratory classroom setting?

Purpose of the Study

The research and findings described in this manuscript represent the fourth phase in two projects that began in 2003. The original project developed a four phase process and SET to evaluate teaching in quantity foods laboratory classes (Chandler, Weber, and Finley, 2007). The current project applies the four phase process to developing a SET to evaluate teaching in a basic food laboratory.

In phase one of the current project, a Delphi panel was organized and through three rounds, developed a SET to evaluate teaching in a basic food laboratory (Chandler, Weber, Finley, and Evans, 2008). Phase two of the

development process consisted of surveying members of the Council for Hotel, Restaurant, and Institutional educators (CHRIE) about the appropriateness of each of the 27 criteria in evaluating teaching effectiveness. Phase three consisted of student evaluation of the preliminary SET. Based on recommendations from faculty and students, a SET consisting of 27 criteria was developed (Weber, Chandler, and Finley, submitted for publication). Phase four is a pilot study of the SET with students enrolled in a basic food laboratory class and is the focus of this study.

REVIEW OF LITERATURE

The literature review covered several subjects pertinent to the development of a student evaluation tool to be used in basic food laboratories. This particular type of laboratory class varies greatly from any standard university lecture course. Current evaluation tools such as the ones required by accrediting agencies are designed for the evaluation of standard lecture classes and are considered unsuitable for evaluation of basic food laboratory courses. In addition, it appears that there have been no studies aimed at collecting student evaluations of teaching effectiveness in basic food laboratory courses.

Student Evaluations of Teaching

University classes are commonly rated using different methods to ensure teaching effectiveness and course content. The most commonly used measure is the student evaluation tool (SET). These SETs are intended to provide valuable feedback to teachers and administrators. They are administered at the end of a course and evaluate only the student perceptions of effective teaching. According to Clayson (1999), SETs are used at four out of five campuses nationwide. There are questions, however, as to whether the results from SETs alone give a one-dimensional limited view of the classroom. The SET is viewed as one-dimensional because it does not include the perspective of other educators (Marsh and Roche, 1997).

Research on developing SETs for hospitality management lecture courses has been reported by Gursoy and Umbreit (2005). The authors of this study of over 25,000 students proposed that student perceptions of the teacher's organization, workload, and instructional abilities influence their perception of learning. The model explained 78% of the variance in the perception of learning.

A study by Simpson and Siguaw (2000) examined faculty perceptions of student evaluations of teaching (SET) and their problems, to ascertain faculty response to SETs and perceptions to the responses, and to consider solutions to perceived problems with SETs. The researchers used a questionnaire to collect the data from a sample of 52 respondents from 15 different states in the United States and other locations outside of the United States. The faculty agreed that there are a number of problems with SETs including the fact that some students may use the SETs as a way of revenge, the question of validity of the evaluation instrument, and the ability of the student to objectively evaluate faculty (Simpson and Siguaw, 2000).

McKeachie (1997), d'Apollonia and Abrami (1997), and Greenwald and Gillmore (1997), all conducted research into the validity of student ratings. These authors stated that faculty use ratings as feedback for how to improve their courses. Their conclusion claimed that student ratings are the most valid data source on teaching effectiveness. However, Marsh and Roche (1997) also stated that multidimensionality in teaching is important not only because of its obvious diagnostic utility as instructor feedback, but also because it provides a more sophisticated and realistic assessment of the various aspects of teaching.

This review shows the strengths of using student ratings as a valuable part of improving teaching effectiveness. It must be recognized that it is the choice of the instructor to incorporate this feedback. Likewise, the review also discusses challenges in using a SET because students can have a limited (one-dimensional) view of the classroom/laboratory. In addition, students may not have enough experience in evaluation techniques to provide quality feedback. The researchers found no studies that attempted to compare educator and student feedback on criteria evaluating teaching effectiveness in foods laboratory courses.

METHODOLOGY

In September 2008, the Principal Investigator established contact with the faculty of several of the leading hospitality management programs to solicit their participation in this phase of the research. Ten universities agreed to participate in the project during the Fall 2008 semester and the Spring 2009 semester. In November 2008, appropriate quantities of the basic foods laboratory student evaluation tool (SET) developed during the first 3 phases of the project were distributed to each of the participating institutions, along with scantron sheets to facilitate data

entry. The evaluation instrument was modified to collect demographic information from the responding students. Faculty at the participating institutions were asked to administer the evaluation instrument before the end of November, and they were provided with self-addressed, postage-paid envelopes for returning the completed instruments to the researchers. Data were received for the Fall 2008 semester from eight universities, and arrangements are in place to replicate this procedure during the Spring 2009 semester.

The survey instrument consisted of the 27 criteria which asked students to rate their instructor on the items according to a 5-point Likert scale that ranged from “never to always,” or from “unacceptable to outstanding,” according to the wording of the criteria, and with 5 being the highest rating and 1 being the lowest rating. The evaluation instrument is available from the authors. Demographic information requested included students’ gender, class level, career aspiration, and grade they expected to earn in the course. Respondents were also asked to compare this evaluation instrument with the one currently being used at their university.

Data analyses were conducted using SPSS (v. 15.0). The first stage of the analysis explored and compared descriptive statistics for the 27 items of the evaluation instrument. The second stage of the analysis considered the content and construct validity and reliability of the instrument. Next, a factor analysis was conducted to reduce a large number of variables to a smaller number of factors, and to describe the relationships among the observed variables.

Findings – Fall 2008

The data for Fall 2008 consisted of 475 useable instruments among the 487 that were received (97.5%) from eight of the ten universities. Twelve responses were deleted because of incomplete evaluation and/or demographic data. Among useable responses, there were 178 male (37.5%) and 297 female respondents (62.5%). Ninety-seven percent of the respondents were in their sophomore (28.6%), junior (41.9%), or senior (26.7%) year of studies, and 94% expected to earn a grade of “B” or higher. Lastly, 66% of the respondents thought the survey was better than the survey sponsored by their university.

The means for the 27 criteria ranged from 4.02 to 4.72 (SD range: 0.646 to 1.019). Table 1 indicates the mean, standard deviation, and 95% confidence interval (lower and upper bound), for each of the 27 criteria of the evaluation instrument.

Table 1
Descriptive Statistics and Reliability (N = 475)

Criteria	Mean	SD	95% Confidence Interval		Alpha
			Lower Bound	Higher Bound	
#1	4.19	.844	4.11	4.26	.961
#2	4.26	.887	4.18	4.34	.961
#3	4.13	.824	4.04	4.21	.961
#4	4.30	.957	4.21	4.39	.961
#5	4.25	.933	4.17	4.34	.961
#6	4.41	.828	4.34	4.49	.961
#7	4.26	.930	4.18	4.34	.960
#8	4.72	.646	4.66	4.78	.961
#9	4.16	.993	4.07	4.25	.961
#10	4.38	.856	4.30	4.45	.960
#11	4.50	.764	4.43	4.57	.961
#12	4.31	.899	4.23	4.39	.961
#13	4.40	.821	4.32	4.47	.960
#14	4.41	.838	4.33	4.49	.960
#15	4.15	1.007	4.06	4.24	.960
#16	4.43	.850	4.35	4.50	.960
#17	4.02	1.019	3.92	4.11	.961
#18	4.38	.844	4.30	4.46	.960
#19	4.55	.785	4.47	4.62	.961
#20	4.36	.782	4.29	4.43	.961

#21	4.31	.817	4.24	4.39	.961
#22	4.44	.778	4.37	4.51	.961
#23	4.54	.755	4.48	4.61	.962
#24	4.50	.716	4.44	4.57	.961
#25	4.36	.835	4.28	4.43	.960
#26	4.20	.858	4.12	4.27	.960
#27	4.09	.981	4.00	4.18	.962

The next aspect of the analysis concerned validity and reliability. The instrument for measuring teaching effectiveness that was developed by a Delphi panel has been reviewed and evaluated by both CHRIE educators and hospitality management students. In both studies, the respondents expressed favorable opinions of all criteria items contained in this instrument. Based on these findings and analysis, it is the consensus of the researchers that the findings of this research project have supported the construct and content validity of the survey instrument (Chandler, Weber, and Finley, 2007).

Reliability was calculated using Cronbach's alpha. To assess internal consistency of continuous scores, a reliability analysis was conducted using SPSS. A coefficient alpha for all 27 variables was produced (see Table 1), the coefficients for all items ranged from .960 to .962, and the overall coefficient was .962. As suggested by Nunnally (1978), 0.7 is an acceptable reliability coefficient, but lower thresholds are sometimes used in the literature.

The next step is to conduct a factor analysis on the 27 criteria items. The data has a fairly normal distribution with skewness ranging from -2.818 (criteria 8) to -.745 (criteria 3). Kurtosis ranged from -.143 (criteria 3) to 9.226 (criteria 8). According to Tabachnick and Fidell (2001), the preferred range is plus/minus 3.0. The kurtosis values above zero indicate a distribution that is too peaked with short, thick tails; the one kurtosis value below zero indicates a flat distribution (Tabachnick and Fidell, 2001). After reviewing the histograms for the 27 criteria items, only three criteria have a kurtosis value outside the acceptable limit (#8, #11, and #23). These items are peaked with responses of "5" on the survey, and are not drastically outside of the normal distribution curve. Upon request, the Principal Investigator will provide the complete statistical analysis to interested individuals.

After this initial analysis of the data set, a factor analysis was performed. An alpha extraction method was chosen to maximize the alpha reliability. A varimax rotation was chosen to minimize the number of high loadings on each factor. The results yielded 3 factors with eigenvalues greater than one, and explaining 62.9% of the variance.

Upon further review of the loading matrix, it is evident that all three factors had loads above .32 (Comrey and Lee, 1992). They suggest loadings in excess of .71 (50% overlapping variance) are excellent; .63 loadings (40% overlapping variance) are very good; .55 loadings (30% overlapping variance) are good; .45 loadings (20% overlapping variance) are fair; and .32 loadings (10% overlapping variance) are poor (Comrey and Lee, 1992). See Table 2 for the loading factors.

Table 2
Factors and SET Criteria (N = 475)

Factor	Criteria Included	Factor Loading
The Instructor	#4- Demonstrated hands-on skills	.692
	#5- Defined structure and objectives	.591
	#6- Ensured appropriate material	.623
	#7- Demonstrated professional rapport	.694
	#8- Dressed professionally	.531
	#9- Provided constructive feedback	.671
	#10- Demonstrated professional behaviors	.748
	#11- Used correct terminology	.631
	#12- Maintained regular office hours	.592
	#13- Utilized professional expertise	.635
	#14- Demonstrated professional body language	.628
	#15- Adapted teaching methods	.655
	#16- Willing to help students with questions	.666

	#17- Used a variety of instruction methods	.584
	#18- Created a positive lab environment	.663
	#19- Maintained a safe workspace	.578
The Students	#20- Followed procedures in syllabus	.653
	#21- Synthesized knowledge	.637
	#22- Actively engaged in lab experiences	.738
	#23- Students dressed professionally for lab	.707
	#24- Used lab time efficiently	.748
	#25- Modeled professional behavior	.744
	#26- Created a positive lab environment	.594
	#27- Prepared in advance for lab	.562
The Course	#1- Outlined the skill competencies	.509
	#2- Challenged students to demonstrate skills	.561
	#3- Objective evaluation of performance	.474

DISCUSSION AND CONCLUSION

The most significant finding of this phase of the research was the high degree of reliability of the evaluation instrument. This was supported by the Cronbach's coefficient alpha, which explained 96% of the variance of the entire evaluation instrument, as well as within each of the 27 criteria. As stated previously, the content and construct validity were well-documented and supported by findings reported from phases two and three of the project.

Within the construct of the evaluation instrument, the factor analysis verified the Delphi panel's grouping of criteria. The criteria were grouped in a manner that allows respondents to evaluate the course, the instructor, and students' participation and involvement in the learning activities. The researchers felt that it was interesting that these 27 criteria loaded into three factors.

The purpose of this study is to conduct a pilot test of the student evaluation tool (SET) that was developed and tested during preceding phases of this ongoing research project. The findings reported in this manuscript are preliminary, in that they reflect the results of the data collected during the Fall 2008 semester and the final set of data will be collected in the Spring 2009 semester. Although the study findings are certainly not conclusive, the findings at this intermediate stage of the pilot test provide strong support for the high degree of reliability of this student evaluation instrument that was designed, developed and tested for application in a basic food laboratory environment.

IMPLICATIONS FOR FUTURE RESEARCH

Much work remains to be done in this research venue. After completing phase four in the Spring 2009 semester, the researchers would like to re-convene the original Delphi panel to consider the appropriateness of a combined evaluation instrument for both the quantity and basic food laboratory environments. In addition to students' evaluation of teaching effectiveness, much work is needed in the venue of peer observation of instruction in laboratory settings. The collective experience of this team of researchers has been that peer evaluation tools currently being utilized are not appropriate for application in a basic or quantity food laboratory setting. This leads to a need for valid and reliable peer evaluation tools for quantity foods and basic foods laboratories.

REFERENCES

- Chandler, J., Weber, M., & Finley, D. (2007). Evaluating teaching effectiveness in a quantity food laboratory setting. *Journal of Hospitality and Tourism Education*, 19(3), 23-31.
- Chandler, J. A., Weber, M., Finley, D. A., & Evans, D. A. (2008). Assessing teaching effectiveness in a basic food laboratory setting: Phase one – Delphi panel study. *Journal of Culinary Science & Technology*, 6(4), 295-307,

- Clayson, D. E. (1999). Students' evaluation of teaching effectiveness: Some implications of stability. *Journal of Marketing Education*, 21(1), 68-75.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4), 743-751.
- Gursoy, D., & Umbreit, W.T. (2005). Exploring students' evaluations of teaching effectiveness: What factors are important? *Journal of Hospitality & Tourism Research*, 29(1), 91-109.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187-1197.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199-213.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn and Bacon.
- Weber, M., Chandler, J.A., & Finley, D. (2008). Evaluating teaching effectiveness in a basic food laboratory setting: Educator and student responses. *Manuscript submitted for publication*.
- Woods, R. H. (2002). *Managing hospitality human resources*. Lansing, MI: Educational Institute of the American Hotel and Lodging Association.
- Yong, Y. W., Keng, K. A., & Leng, T. L. (1988). A delphi forecast for the Singapore tourism industry: Future scenario and marketing implications. *International Marketing Review*, 6(3), 35-46.