A New Paradigm in Student Course Evaluation: From Instructor Satisfaction to Course Content

Oscar T McKnight, Ashland University
Ronald Paugh, Ashland University
M Manzo
A NEW PARADIGM IN STUDENT COURSE EVALUATION: FROM INSTRUCTOR SATISFACTION TO COURSE CONTENT

Oscar McKnight, Ashland University
Ronald Paugh, Ashland University
Maria Manzo, Ashland University

ABSTRACT

Historically, student course evaluations have been viewed as “satisfaction measures” reporting what students “like” or “dislike”. This exploratory research examines the feasibility of redefining student course evaluations as a tool for curriculum development using a factor analytic solution. Research results suggest that a paradigm shift is possible.

INTRODUCTION

If one were to conduct a thorough review of the existing research on written and objective student evaluations as they relate to the teaching performance of university instructors, a rich research tradition would be discovered. In a “brief” review of Wachtel’s (1998) historical outline, student evaluations were conducted in the 1920’s, but it was not until the 1970’s that the “golden age” of research on student evaluations was realized. Prior to this, evaluations were mostly hand written, and non-standardized. As a result, design and content were weak in terms of statistical rigor and in need of revision.

However, standardization of instruments was not universally viewed as a positive development. Powell (1978) summarized many professionals’ sentiments by stating that most of the research supporting the validity of student evaluations came from those publishers or authors who developed the scale and offered their product for a price. This sentiment presupposes that standardized student course evaluation forms are constructed for the sole purpose of making money, not for providing accurate and useful information related to the course or instructor.

This less than positive view of student course evaluations is held by many authors, instructors, and educators who assert that student course evaluation models reflect uncertainty at best (see Franklin and Theall, 1989). However, for a balanced review of the literature pertaining to student course evaluations, the interested reader should consult Wachtel (1998). Wachtel provides a historical overview, and advocates the professional necessity of developing multiple approaches to evaluation. The purpose of this introduction, however, is not to establish a “pro” or “con” position on the use of student course evaluations, but rather to stress the necessity of continued research and to acknowledge that student satisfaction measures can be controversial.

LITERATURE REVIEW

In 1978, Feldman addressed course characteristics and how college students rated their teachers. The focus was on the relationships between course context and ratings of instruction. The researcher found weak associations, but stressed that clear patterns do emerge. For example, there was an inverse relationship between class size and class ratings; ratings were higher for upper division courses or electives; and ratings for teachers in the humanities, fine arts and languages tended to be higher than other disciplines. However, Feldman did introduce in discussion, and literature review, that “course grade” or “expected course grade” was a strong predictor of course ratings, and that in most studies, this variable should be either held constant or acknowledged as a factor. Feldman’s research raised a cautionary note when analyzing student course evaluation data.

A study published by Naftulin, Ware, and Donnelly (1973) dubbed “The Dr. Fox Effect” and follow-up studies (e.g., Ware and Williams, 1979; Marsh and Ware, 1982) examined the phenomenon of enthusiastic teaching and its effect on student
evaluations in the absence of adequate content. The classic "Dr. Fox" study involved the use of a professional actor to deliver a lecture. Later, the researchers administered and assessed student evaluations. Although the original study was recognized as having many methodological problems, the impact of style over substance entered the "satisfaction" discussion.

Marsh (1983) began a research project by emphatically stating that most research fails to recognize the multidimensionality of student ratings, and introduced the potential for researcher biases. With this in mind, Marsh decided to examine six groups on sixteen student-class-instructor characteristics, using stepwise multiple-regression. The results indicated higher student ratings with students who had a prior interest, expected a higher grade, increased workload difficulty, and had taken the course as a general interest. Overall, Marsh concluded that some variables will affect teaching effectiveness and will ultimately be reflected in student ratings. However, Marsh never addressed if satisfaction ratings influence course development. Yet, it is not uncommon to find research that examines student satisfaction in relationship to faculty assessment with an implicit assumption that it improves course content.

Cranton and Smith (1986) opened their research on the effect of course characteristics on student ratings of instruction by pronouncing that "student ratings of instruction are the major, if not the only, component of the evaluation process at most universities in North America" (p. 117). They warn that researchers and administrators should interpret results with caution and be wary of the instructional setting in which the ratings were collected. Specifically, the researchers felt that students will adjust their ratings to account for the instructional setting. The overriding implication of this research was not to over interpret, generalize, or evaluate faculty solely on student satisfaction surveys or course evaluations.

Wigington, Tolleson and Rodriguez (1989) reiterated the professional plea not to use student ratings as a measure of teaching expertise. Their research findings indicated that class level or size, type of instruction, instructor reputation, rank, and sex, all appear to influence ratings. The researchers concurred with those earlier warnings regarding the interpretation of student course evaluations. Moreover, they strongly affirmed that evaluating courses or faculty by looking at composite mean scores to determine which instructors are "better" is ill-advised, given the evaluation conditions and circumstances.

RESEARCH OBJECTIVES

Civian and Brennan (1996) remind us of the "wealth of research" that exists in the area of student course evaluations. Studies range from analyses of validity and reliability to student and course characteristics and their effect on overall ratings. However, merging an evaluation method with course construction and content has not been given nearly as much attention in the professional literature.

This exploratory research examines the fundamental belief that feedback from student ratings can help to improve instruction. This predominant theme reverberates through the work of Overall and Marsh, 1979; Cohen, 1980; Menges, 1991; and Marsh and Roche, 1993. There are also many controversies surrounding student evaluations of teacher effectiveness since there has never been agreement on the appropriate criteria for measuring teacher effectiveness (Marks, 2000). This is the underlying sentiment of Wigington, Tolleson and Rodriguez when they assert that evaluating faculty by a "mean" score is inappropriate. Moreover, it is critical to this current research to reiterate the key finding by Marsh (1982), and later corroborated by Cashin (1988) — the instructor, not the course, is the primary determinant of student ratings; however, what a high or low satisfaction rating really indicates is patently speculative.

A PROPOSED PARADIGMATIC SHIFT

This study proposes to assess student satisfaction using a new paradigm. Specifically, it is a procedure that employs a factor analytic solution on the total number of questions/responses asked/received from a student with respect to satisfaction and then designates a "marker item" within the administered questionnaire. This marker item could be a specific question related to course content, student recommendation, student retention, or any other issue. However, the researcher's objective would be to determine where this marker item loads on the factor matrix. Factor analysis as an approach is widely used in evaluating student ratings (Linn, Centra, and Tucker, 1975; Marsh and Hocevar, 1983; Jackson, 1999). Nevertheless, employing a specific marker item with theoretical implications is a unique application. Therefore, what follows should be considered an illustration of a tool for curriculum development that details how an institutional/department researcher could assess
student satisfaction with respect to course content from one semester or academic year to the next.

A TOOL FOR CURRICULUM DEVELOPMENT

This research examines the relevance of a student recommending a course to another student. This “marker item” is conceptually related to both student satisfaction and student retention. Therefore, this investigation examines three questions relevant to the professional literature and offers an exploratory model that has the potential to become a new evaluation paradigm for the use and interpretation of student satisfaction surveys. In the example that follows, the questions, data, and interpretations that are introduced have practical implications. The questionnaire used was developed specifically for the class under investigation. Therefore, the questionnaire, its construction, and administration must be viewed as a working model that demonstrates a potential paradigmatic shift in the use of student evaluations. The specific questionnaire items and components are secondary to the process and interpretation of the research findings. The following three questions are stated as hypotheses to illustrate how this process could be used as a tool for curriculum development.

HYPOTHESES

H1: Can student ratings be used to validate course content by employing a factor analytic solution when introducing a marker item?

H2: Can a factor score representing course content account for a significant amount of variance when predicting if the student will recommend the course to other students?

H3: Can a mean score (average) of student ratings of instructor satisfaction account for a significant amount of variance when predicting if the student will recommend the course to other students?

METHODOLOGY

The population chosen for this exploratory research was 372 freshmen enrolled in 29 Introduction to University Life courses at a small, residential, private university located in the midwest. The rationale for choosing this population and course were simple and pragmatic: the course had a curriculum with content, not just social activities to evaluate; it was a suggested course for all entering freshmen; and the students had an opportunity to rate their satisfaction with the instructor, advisor, campus community, course content, and other general areas.

All questionnaire items used a Likert scale, assigning a number 1, meaning “strongly agree” to 5, meaning “strongly disagree” with a sentiment expressed in each question. For example, one item expressed the sentiment, “The instructor encouraged me to develop my desire to learn”. All other questions had a similar construction. For clarification, the questionnaire is not presented in its entirety for two pragmatic reasons. First, it is not a standardized instrument, as it was developed specifically for this university class. Second, and perhaps most important, copyright law prohibits the publication of the instrument. However, the factor labels and variance accounted for were based on the actual rotated matrix.

RESULTS

Statistically significant results were found in each of the three stated hypotheses (H1; H2; and H3). Since H2 was dependent on the findings in H1, the factor analysis was performed first. The factor analysis results, utilizing a varimax rotation, yielded a five-factor solution. When combined, the five factors accounted for 59.7 percent of the variance. However, factor one independently accounted for 34 percent of the variance.

Conceptually, factor one included items that were associated with course content. The marker item [I would recommend the course to other students] loaded on this factor. Factor two was associated with instructor satisfaction. Factor three represented satisfaction with social involvement. Factor four was a conceptual measure of advising. Factor five was dominated with items related to career advising services. There was a sixth factor, but it was eliminated since it accounted for less than two percent of the variance and exhibited no clean item or loading above .50.

Table 1 highlights the results for H2. The factor score representing “course content” (factor one) accounts for a significant amount of variance (43.0%) when predicting if a student would recommend the course to other students. Therefore,
findings suggest that "course content" does predict if a student recommends the course to others.

The results of H3 are listed in Table 1. H3 accounts for a significant amount of variance (17.0%)

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>R-square</th>
<th>Adjusted R-square</th>
<th>t</th>
<th>Alpha</th>
<th>Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2</td>
<td>.673</td>
<td>.453</td>
<td>.43</td>
<td>19.89</td>
<td>.05</td>
<td>.0001</td>
</tr>
<tr>
<td>H3</td>
<td>.423</td>
<td>.179</td>
<td>.17</td>
<td>72.68</td>
<td>.05</td>
<td>.0001</td>
</tr>
</tbody>
</table>

Note: Total N size was 340; results from 32 questionnaires were not included in analysis because of response set bias; for example, answering all questions (n=35) with the same response pattern.

### DISCUSSION

Although the questionnaire used in this exploratory research exemplifies a paradigmatic shift in process, research findings indicate several interesting points for discussion. For example, the five factors extracted while investigating H1 are indicative of a valid course framework, exhibiting logical loadings. One would hope to find overall instructor ratings falling on various aspects of "Instructor Satisfaction" versus some other "characteristic" like Career Services, Advising Services, or Course Content. This was found in the factor matrix. Findings support the previous researchers' assertions (Overall and Marsh, 1979; Menges, 1991; and Marsh and Roche, 1993) that student feedback in the form of satisfaction ratings can enhance learning objectives through instruction. Therefore, using student satisfaction ratings to determine or influence course content appears warranted.

In H2, the research question addresses the impact that course content, as a factor score, has on securing a student recommendation. Findings suggest that course content accounts for 43% of the variance when predicting a recommendation. Therefore, areas related to course content are a strong indicator of receiving a student recommendation for others to take the course.

In H3, students' satisfaction with the instructor accounted for 17% of the variance when predicting if a student would recommend the course in solution. Hence, student rating of "instructor satisfaction" does predict better than chance that a student will recommend the course to other students.

It would appear that the results from H2 and H3 are contradictory, but the findings are logical. The results suggest that both course content and faculty satisfaction will predict if a student recommends the course to another student. This should be the goal in higher education, to have not only a quality instructor but also satisfying course content. It is in the best interest of both the institution and student to view satisfaction as a multidimensional construct (see Marsh, 1983) and not as a "black or white", "either-or", proposition. It is possible to have both, and this exploratory research supports such an assertion. Therefore, this research supports the caution of Wigington, Tollefson and Rodriguez (1989) who stated that evaluating faculty by looking at a composite mean (average) student satisfaction score in order to determine which instructors are "better" — is misguided. However, the results do suggest that some areas of course content may be more important to securing a "student recommendation" than others, supporting the work of Feldman (1978) and the paradigm offered.
The current researchers would be remiss if they did not discuss other possible reasons for the widespread use of the student course evaluation or satisfaction survey. First, it is a tangible indicator that can be used to demonstrate everything from upholding the university mission to appeasing accreditation agencies. Second, student course evaluations can be used as a marketing and enrollment management tool, since student satisfaction can be linked conceptually with retention issues (Elliott and Shin, 1999). Finally, indicators of student satisfaction are often times a critical criterion in grant reports and other funding requests.

LIMITATIONS AND FUTURE RESEARCH

This study examined the results of a student satisfaction questionnaire administered at one midwestern university. Since every university is defined by a myriad of characteristics like type, location, and selectivity, caution must be exercised when generalizing the results (see Cranton and Smith, 1986). For example, this university population exhibits average ACT and SAT admission scores. Therefore, extending results to a more or less selective university population may be unwarranted, but an area worthy of further investigation.

It would have been desirable to introduce the complete questionnaire, its items, and internal validity studies. Since the authors did not have this option, it was decided to present the results “as if” it were a standardized instrument. However, given the pragmatic nature of student evaluations in many universities and colleges, this should not be considered an unusual practice.

Moreover, although this exploratory research successfully employed factor analysis to determine the relationship between receiving a student recommendation and core class components, the use of marker items representing “complete satisfaction” should be examined further in subsequent research. It would be useful to study the impact on course development using “student satisfaction” measures as a means by which to validate course content, and, as a continuous improvement initiative. For example, by applying this process to multiple courses or perhaps an entire marketing program, a professional body of literature in terms of program continuity could be developed and utilized.

It is important to reiterate what Wachtel (1998) suggested as a worthy endeavor — conduct further research on faculty members to determine how often and what types of changes they make in their pedagogical approach based on the results of student course evaluations. This proposed paradigm used a factor analytic solution. However, factor analysis is really a descriptive tool. The real paradigm shift materializes when one employs the factor results from one semester to another, using an experimental or predictive model based on earlier descriptive findings. Only then can a new paradigm in student course evaluation emerge — one that specifically focuses on course content.

CONCLUSIONS

This study began by introducing the controversial nature of student satisfaction surveys and the multidimensional aspects of construction and interpretation. The current authors did not take a position on the use of satisfaction surveys, but rather offered another way to supplement and interpret student satisfaction feedback in terms of course content. Although this research examined student satisfaction as it related to recommending the course to others, a host of marker items could have been introduced. But this is the essence of the new paradigm offered. Simply stated, a researcher has the ability to designate questions (i.e. marker items) related to: retention, majors, minors, certifications, employment, internships, or any other issue, with any standard or “non-standard” instrument. The paradigm shift derives from using this information (i.e. through a factor analytic solution) the next semester (or academic year) to evaluate or predict results. When this approach is taken in a disciplined manner, college or university officials have a new curriculum tool that has predictive and descriptive capabilities. Therefore, a new paradigm emerges, shifting institutional researchers away from instructor satisfaction, and toward course and curriculum development.

REFERENCES


Naftulin, Donald H., John E. Ware and Frank A. Donnelly (1973), "The Dr. Fox lecture: A paradigm of educational seduction," Journal of Medical Education, 48, 630-635.


Ware, John E. and Reed G. Williams (1979), “Seeing Through the Dr. Fox Effect: A Response to Frey,” Instructional Evaluation, 3, 6-10.