The Results of Student Ratings: Paper vs. Online

Te-Sheng Chang
National Hualien Teachers College
Graduate Institute of Compulsory Education

Abstract

The purpose of this study was to compare the results of student ratings of their instructors via paper and online surveys. The sample consisted of students at 624 undergraduate courses at National Hualien Teachers College in the fall semester of 2001: 198 (31.73%) freshman courses, 161 (25.80%) sophomore courses, 146 (23.40%) junior courses, and 119 (19.07%) senior courses. The instrument was the Students’ Rating of Instructors (SRI) form developed in 1995 at this college. The SRI form was composed of 13 questions rated on a 5-point Likert scale, ranging from “strongly agree” (5 points) to “strongly disagree” (1 point). These 13 items were clustered around four teaching factors: Preparation/Planning, Material/Content, Method/Skill, and Assignments/Examination. The scores on these four factors were added to give the total score (rating) for a faculty member.

The paper scores are significantly higher than the online scores for all of the evaluation items. There are 573 (91.8%) courses for which the average total paper evaluation score is higher than the average total online evaluation score, but only 51 (8.2%) courses for which the average total online score is higher than the average total paper score. These results indicate that the majority of students in all courses give the instructors a higher score when they evaluate them using the more traditional method: sitting in the classroom and using paper forms.

Key words: Student Rating of Instruction, Paper Survey, Online Survey

Introduction

Student ratings of instruction are widely used to evaluate teaching effectiveness in higher education (Seldin, 1999). In the United States, for example, Wagenaar (1995) stated that well over 90 percent of schools currently use student ratings for assessing the teaching staff. Wilson (1998) predicted most colleges would consider student ratings of instruction as a measure of "teaching quality" for their faculty. Student ratings of instruction are ongoing and an ubiquitous rite on college and university campuses throughout the United States and around the world.

In a comprehensive review of research concerning student ratings of college and university instructors, Wachtel (1998) noted that “after nearly seven decades of research on the use of student evaluations of teaching effectiveness, it can safely be stated that the majority of researchers believe that student ratings are a valid, reliable, and worthwhile means of evaluating teaching” (p. 2). Yet, there are many others who would argue against this point, as more recent research has indicated that
methodological factors and situational characteristics can affect the validity of student evaluations. For example, it has been shown that student ratings may be affected by student perceptions of grade leniency (Nimmer & Stone, 1991), by instructor enthusiasm (Williams & Ceci, 1997), and by survey procedures (Layne, DeCristoforo, & McGinty, 1999).

Traditionally, student ratings are usually administered in class using paper questionnaires - a costly, time-consuming process that is inconvenient for faculty and often restricts the thoughtfulness and the depth of student responses (Johnson, 2002). The increasing use of technology in education, especially the World Wide Web, has led to the possibility of online administration and reporting of student ratings. In a recent survey of the 200 most wired colleges in the United States, 25% of respondents said they were already using or plan to covert to online student ratings (Hmieleski, 2000). A web search of institutions using online evaluations yielded over 80 universities converting paper into online student ratings for some courses (e.g., specific colleges or departments, online courses) and three universities using online ratings for all courses on campus (Johnson, 2002). Like the universities in the United States, some colleges and universities in Taiwan have used online ratings systems for all courses on campus. In a recent survey of 76 colleges by Chang (2001), ten schools completed the online ratings systems for all courses on campus and nine schools were planning to convert paper ratings systems to online systems (Chang, 2001). For example, National Hualien Teacher College has moved all of its course evaluation surveys to a web-based system since the fall semester of 1998.

Administering an online survey rather than a paper one shows a number of advantages. For data collection, there are four advantages: (1) survey data entry is automated, (2) missing data can be eliminated, (3) there are no out-of-range responses, and (4) complex item branching, transparent to the respondent, can be used (Rosenfeld & Stephanie, 1993). When surveys are administered via computer or online, data entry takes place as the respondent completes the survey. This saves time and leads to more accurate data because one of the stages at which errors are normally introduced into the database is skipped. In general, survey data collected via a computer or online will not contain missing responses. Besides, online surveys have many economic benefits, including reduced processing time and costs, ease of administration, and more detailed, faster reports (Johnson, 2002; Llewellyn, 2002).

Though the online survey method offers a number of advantages over the paper method, little is known about the extent to which its results are comparable to those obtained through the traditional mode, the paper method. Without answering this question, online student ratings program’s accountability demands cannot be met. Therefore, the major purpose of this study is to compare the results of student ratings via paper and those by online surveys. The research questions guiding the study are as follows.

1. Is there any significant difference in average ratings comparing paper survey to online survey method?
2. Do classes give courses the same ratings scores across paper and online survey methods?
Literature Review

Student ratings of instruction have undergone several purposes. Marsh and Roche (1993) cited three general reasons for evaluation of teaching by students: (1) to provide formative feedback to faculty for improving teaching, course content and structure; (2) to provide a summative evaluation of teaching effectiveness for promotion and tenure decisions; and (3) to provide information to students for the selection of courses and teachers. As Wilson (1998) predicted, most colleges use student ratings of instruction as a part of teaching quality for their faculty promotion. Since summative evaluation of instruction has profound implication for faculty with respect to promotion, tenure, and merit pay increases, care should be exercised in the development of any rating instrument and the procedure for administering this instrument.

In the development of student ratings instruments, many studies have examined issues such as the development and validity of an evaluation instrument (e.g., Marsh, 1987), the validity (e.g., Cohen, 1981), and reliability (e.g., Feldman, 1977) of student ratings in measuring teaching effectiveness, and the potential bias of student ratings (e.g., Centra & Gaubatz, 2000; Chang, 2000; Feldman, 1993).

In the procedure for administering the ratings instrument, the paper survey is the traditional and most common method for collecting data in organizations. Because of their longer history and more frequent use, paper surveys are usually considered as the standard against other forms of survey administration, such as face-to-face interviews, telephone surveys, or electronic surveys. Both face-to-face interview surveys and telephone surveys have a number of disadvantages related to paper and computer surveys. They are expensive to conduct, both because of the costs of training interviewers and because of the hours needed to gather data is very time consuming. If interviewers are not trained or are poorly trained, the quality of data may be variable (Miller, 1991).

Paper surveys have a number of important strengths: They are easy and efficient to administer, inexpensive, and familiar to those being surveyed. It is easy to establish confidentiality on a paper survey. Paper surveys, however, have a number of drawbacks. Individuals completing paper surveys may skip items or choose multiple responses (Doherty & Thomas, 1986). Also, translating the survey responses from paper to the computer for analysis is time-consuming and may introduce additional errors into the data base (Rosenfeld & Booth-Kewley, 1993).

The major weakness of student ratings via paper survey is the difficulty in ensuring the integrity of the survey procedures. With the increasing use of student ratings for personnel-decisions reference, the lack of survey administration standardization procedures is very crucial and particularly troubling (Layne, DeCristoforo, & McGinty, 1999). Ory (1990) indicated that his campus had reported violations of survey administration integrity “due to instructors administering the forms as they walked around their classrooms or collecting and reading the evaluations before sending them to the campus.
There are some problems in Layne, Decristoforo and McGinty (1999) and Hardy (2002) studies. First, the sampling error occurs due to different individual bias in Layne, Decristoforo and McGinty (1999) study. According to this study, students in the classes were assigned either the tradition survey group or the electronic group. That means the different student groups using different survey methods. The effect of survey methods was mixed with the bias of the different individuals. Second, the inferences of the previous research findings are limited since the sample sizes are quiet small. There are 66 classes in Layne, Decristoforo, and McGinty’s study and only 31 in Hardy’s study. Third, the response rates for online ratings in Layne, Decristoforo and McGinty (1999) study were not high. The total response rate for the electronic group was 47.8% and 60.6% for the traditional survey group.

This study at National Hualien Teachers College differed from the previously mentioned in three ways. First, instead of using different raters (different students) in the previous studies, this study asked all participants (same students for the same instructor and same course) to respond to both survey methods. That is, this study took a dependent-sample design to reduce sampling error due to different individual bias. Second, the sample size is much larger than those of the previous research. This study consisted of 624 courses over a college campus. Third, the response rate for online ratings in this study was high because online evaluation was a requirement for all participants before they can become registered for the following semester.
Method

Data

The sample consisted of 624 undergraduate courses with 198 (31.73%) freshman courses, 161 (25.80%) sophomore courses, 146 (23.40%) junior courses, and 119 (19.07%) senior courses from National Hualien Teachers College in Taiwan in the fall semester of 2001. At the time of the study, the total courses offered by the school were approximately 1052. The students in all the courses were asked to do the evaluation on both paper and online methods. A high response rate was achieved, 78.99% (831/1052=78.99%) for paper survey and 95.25% (1002/1052= 95.25%) for online system, respectively. Both data files were matched by teacher identification numbers and course section codes. The matched data included 793 courses. For the representative reason, the class size less than 5 were deleted from study. The final data file consisted of 624 courses ranging from 5 to 51 students.

Measure

The instrument used in the study was the Student Ratings of Instruction (SRI) form developed in 1995 at the college. The rating form was composed of 13 questions rated on a 5-point Likert scale ranging from strongly agree (5-point) to strongly disagree (1-point). These 13 items were clustered around four teaching effectiveness factors: Preparation/Planning (Item 1, 2, 3), Material/Content (Item 4, 5, 6), Method/Skill (Item 7, 8, 9, 10), and Assignment/Examination (Item 11, 12, 13). The sum of these four factors was considered as a total rating score for a faculty member. A panel of 15 evaluation experts for content validity validated all items.

The online survey mirrored the paper version in content. The question section consisted of the three basic information items and the 13 ratings items used for data analysis in this study. In addition to the question section of the survey, both versions included another two sections, one having five items for student self-evaluation and the other having one item for open-ended comment. Thus the total number of items on the SRI administered in this study was 19. Since this study was focused on the student ratings of instruction, the data from these two sections were not analyzed.

Survey Administration

During the last four weeks of the semester 2001, all the instructors on campus were sent an invitation letter that explained the purpose of the study and the mechanics of the survey. Each instructor was asked to designate an individual (e.g. head of the class) to administer the survey. All the survey administrators attended a half-hour training session, during which they were provided with materials for the traditional survey and a list of all the courses in the department.

Because the course opinion survey had already been converted to an online system for the school from the fall semester of 1998, only the paper evaluation form, with an explanation cover letter, was mailed to all the courses in the last two weeks of the semester. Approximately 10 minutes before the end of the last class in each course, the instructor left the room, and the survey administrator explained to the entire class that a simple comparison was to be made between the
paper survey and the online survey. All of the students completed both the paper and the online survey. They completed paper evaluation form together in the classroom setting. Students completed online evaluation survey outside the classroom, either before or after they completed the paper survey.

Students in both survey formats completed the surveys under conditions of complete anonymity and there was no coding scheme employed. Although students in the online format had to use their identity number to gain access to the system, ratings results were stored in a separate database with no identifying information.

Analytic Strategy

All analyses were performed on class-average responses for the sample. The preliminary principal components analysis and the $\alpha$ coefficient of internal consistency reliability on the 13 items were performed separately for the paper and the online survey methods. The correlation coefficients between these two factor patterns were computed by using the Pearson correlation formula in order to determine the degree of similarity between the underlying factor structures of the paper and the online survey groups.

Descriptive data was provided to assist with the interpretation of the differences between the paper and the online survey results. The mean and standard deviation for the paper and the online survey methods on each item, each factor, and the total ratings of the SRI questionnaire were accordingly computed. To investigate whether the ratings were influenced by the survey method, each ratings item, four factors and the total score were analyzed using a dependent $t$ test (same instructors, same courses, and same raters but different survey modes).

In addition, the scores differences between these two methods were computed by subtracting the online scores from the paper scores for each of the 624 courses. The courses scoring above zero were classified as paper-higher courses. Those scoring below zero were classified as online-higher courses. Those scoring equal to zero were classified as paper-online-equal courses. The percentage for each category was computed for each level and for all of the courses.

Results and Discussion

Before the primary research questions were addressed, preliminary factor analyses on the 13 core teaching items were performed separately for the paper and the online survey methods. The factor loadings for the paper evaluation items designed to measure each factor are large, between .850 and .937. The four factors account for 91.25% of the total variance on the paper evaluation. Similarly, the factor loadings for the online evaluation items designed to measure each factor are consistently large, between .864 and .950. The four factors account for 92.28% of the total variance on the online evaluation. The $\alpha$ coefficients of internal consistency reliability are .9792 and .9805 for the paper and the online evaluation scores, respectively. The factor loadings and the $\alpha$ coefficients confirm that the SRI is a valid and reliable instrument regardless of the survey administration methods. The correlation coefficients for the two factor patterns
are.790,.839,.837,.790, and.834 for Preparation/Planning, Material/Content, Method/Skill, Assignment/Examination, and the total, respectively.

Table 1 lays out the mean and standard deviation for the paper and the online evaluation scores and the mean differences between these two scores for each evaluation item of the SRI. The ranges are 3.966 to 4.219 for the paper survey and 3.711 to 3.986 for the online survey. The highest scored item for both survey methods is Item 4, "relates the material of this course with other areas of knowledge". The lowest scored item for both survey methods is Item 12, "gives fair grades".

It seems that most students give the instructors the above-average evaluation score. Why is there a large cluster of score above the average? One of the possible reasons is positive-biasing effect. Researchers in the general survey research area have recognized the positive-biasing effect that self-report surveys have had on survey responses. This positive-biasing effect exists when respondents exhibit a tendency to deny undesirable characteristics and instead respond in a socially desirable manner (Phillips & Clancy, 1970). Likert-type scales, the most frequent scale applied in student ratings of instruction (Berk, 1979), are particularly vulnerable to this yea-saying or acquiescence response set (Couch & Keniston, 1960).

In fact, Centra (1979) summarized a 1975 educational testing service study. He did research on student response tendencies of over 400,000 faculty members, and determined that only 12% of this group received below average ratings due to positive-response bias. Arreola and Aleamoni (1990) emphasized, the research on student ratings has showed an obvious positive-response bias, which needs to be taken into account in interpreting and using results. They advocated the use of normative data to help counteract this rater leniency effect.

Previous research studies (e.g. DeCristoforo, 1992; Sproull & Kiesler, 1991) have revealed that students completing electronic surveys may feel more anonymous and thus become more involved with the nature of the electronic survey method. They may also respond more honestly than in responding with other survey techniques, such as the paper survey. It was initially assumed that students who completed online surveys might become so involved with the interactive online survey method that their responses may negate this positive-response bias. This obviously does not occur in this study since survey results do not vary according to survey method.

Based on Table 1, the paper scores are significantly higher than the online scores on each item. The ranges of the t values are in the 19.028 - 29.411 range for individual rating items. The p values of all the dependent t tests are less than .001. That is, paper evaluation scores are significantly higher than online evaluation scores on each item.

The summary of dependent t tests between these two scores for each factor and total scores is showed in Table 2. The ranges of the evaluation scores are in the 3.980 - 4.095 range for the paper survey and in the 3.741 - 3.851 range for the online survey. The highest scored factor for both survey methods is Material/Content. The lowest scored factor for both survey methods is Method/Skill.
Table 1. The Summary of Dependent t Test for the Survey Method on Each Ratings Item (N=624)

<table>
<thead>
<tr>
<th>Dimension/</th>
<th>Rating Items</th>
<th>Paper</th>
<th>Online</th>
<th>Mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation/Planning</td>
<td>1. is concerned about the effectiveness of his/her teaching</td>
<td>4.028</td>
<td>3.8364</td>
<td>.47</td>
<td>19.028***</td>
</tr>
<tr>
<td></td>
<td>2. provides a detailed course syllabus</td>
<td>4.018</td>
<td>3.811</td>
<td>.373</td>
<td>20.371***</td>
</tr>
<tr>
<td></td>
<td>3. states course objectives for each class section</td>
<td>4.019</td>
<td>3.800</td>
<td>.365</td>
<td>22.136***</td>
</tr>
<tr>
<td>Material/Content</td>
<td>4. relates the material of this course with other areas of knowledge</td>
<td>4.219</td>
<td>3.986</td>
<td>.332</td>
<td>25.152***</td>
</tr>
<tr>
<td></td>
<td>5. demonstrates knowledge and makes it clear how each topic fits into the course</td>
<td>4.146</td>
<td>3.888</td>
<td>.398</td>
<td>28.158***</td>
</tr>
<tr>
<td></td>
<td>6. is aware when students are having difficulty in understanding a topic</td>
<td>3.977</td>
<td>3.733</td>
<td>.367</td>
<td>23.994***</td>
</tr>
<tr>
<td>Method/Skill</td>
<td>7. establishes and maintains an interaction</td>
<td>4.036</td>
<td>3.794</td>
<td>.442</td>
<td>24.376***</td>
</tr>
<tr>
<td></td>
<td>8. keeps the course moving rapidly enough for the material</td>
<td>3.993</td>
<td>3.753</td>
<td>.346</td>
<td>25.575***</td>
</tr>
<tr>
<td></td>
<td>9. explains material clearly</td>
<td>3.981</td>
<td>3.733</td>
<td>.393</td>
<td>24.760***</td>
</tr>
<tr>
<td></td>
<td>10. is helpful with difficulties</td>
<td>3.967</td>
<td>3.738</td>
<td>.408</td>
<td>23.558***</td>
</tr>
<tr>
<td>Assignment/Examination</td>
<td>11. gives good comments on written work</td>
<td>3.987</td>
<td>3.747</td>
<td>.380</td>
<td>21.913***</td>
</tr>
<tr>
<td></td>
<td>12. gives fair grades</td>
<td>3.966</td>
<td>3.711</td>
<td>.324</td>
<td>26.615***</td>
</tr>
<tr>
<td></td>
<td>13. gives exams and papers appropriate for the course</td>
<td>4.069</td>
<td>3.814</td>
<td>.305</td>
<td>29.411***</td>
</tr>
</tbody>
</table>

Note. *** p < .001

From Table 2, the paper scores are significantly higher than the online scores on each of the four factors and the total evaluation scores. The t values are 23.693, 30.011, 27.563, 28.957, and 30.270 for Preparation/Planning, Material/Content, Method/skill, Assignment/Exam, and Total, respectively. The p values of all these dependent t tests are less than 0.001. Like the findings in Table 1, the findings in Table 2 indicate that paper evaluation scores are significantly higher than online evaluation scores.

Table 3 shows the number and percentage for course levels on student ratings by survey methods. When all courses are taken into account, there are 573 (91.8%) courses being higher for paper scores, 51 (8.2%) courses being higher for online scores, and 0 (.0%) courses being equal for the total evaluation. Similarly, there are 536 (85.9%), 568 (91.0%), 554 (88.8%), and 564 (90.4%) courses being higher for paper scores for Preparation/Planning, Material/Content, Method/Skill, and Assignment/Examination, respectively.
Table 2. *The Summary of Dependent t Test for the Survey Method on Each Factor and Total scores*  
\((N=624)\)

<table>
<thead>
<tr>
<th>Evaluation Dimension</th>
<th>Paper</th>
<th>Online</th>
<th>Mean diff</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
<td>(SD)</td>
</tr>
<tr>
<td>Preparation/Planning</td>
<td>4.023</td>
<td>.334</td>
<td>3.815</td>
<td>.338</td>
</tr>
<tr>
<td>Material/Content</td>
<td>4.095</td>
<td>.361</td>
<td>3.851</td>
<td>.356</td>
</tr>
<tr>
<td>Method/Skill</td>
<td>3.980</td>
<td>.388</td>
<td>3.741</td>
<td>.365</td>
</tr>
<tr>
<td>Assignment/Exam</td>
<td>4.008</td>
<td>.341</td>
<td>3.758</td>
<td>.322</td>
</tr>
<tr>
<td>Total</td>
<td>4.031</td>
<td>.341</td>
<td>3.796</td>
<td>.333</td>
</tr>
</tbody>
</table>

Note. *** \(p < .001\)

In terms of the total score, there are 170 (85.9%), 152 (94.4%), 140 (95.9%), and 111 (93.9%) courses being higher for paper scores for freshman, sophomore, junior, and senior courses, respectively. Although the percentages of paper-higher courses are much more than those of online-higher courses for all different levels, the percentage of paper-higher courses for the freshman level is relatively lower than those of the other three course levels. For freshman courses, there are 159 (80.3%), 167 (84.3%), 159 (80.3%), 168 (84.8%), and 170 (85.9%) courses being higher for paper scores for Preparation/Planning, Material/Content, Method/Skill, Assignment/Examination, and the total, respectively. For the rest of the level courses, the percentages are higher than 90.0% for most of the evaluation factors and total score. These results indicate the majority of courses give the instructors higher scores when they do the faculty teaching evaluation with a traditionally paper-pencil way in the classroom.

Why are the paper evaluation scores significantly higher than the online evaluation scores? Does this mean instructors rated by paper format teach more effectively than instructors rated by online format? The answer is negative, since the instructors are rated by the same students with both paper and online survey methods. One of the main reasons that paper evaluation scores are significantly higher than online evaluation scores is due to the way the survey is administered. The findings indicate that it is more probable for students to give instructors higher evaluation scores when they complete the rating form by a paper survey.

It seems that most students felt the paper method afforded a lower degree of anonymity than the online method. One of the possible reasons is that students feel more secure and free to write the honest truth since they do not feel as though their instructor would find out what they had written about him/her. The other reason is that online survey was administered without any involvement of the instructor (Layne, DeCristoforo, & McGinty, 1999). Previous research related to electronic survey technology revealed that students who used the electronic method gave fewer socially desirable responses (Erdman, Klein, & Greist, 1983; Kiesler & Sproull; Martin & Nagao, 1989).
Table 3. The Number and Percentage for Course Level on Student Ratings by Survey Methods

<table>
<thead>
<tr>
<th>Course level</th>
<th>Freshman (N=198)</th>
<th>Sophomore (N=161)</th>
<th>Junior (N=146)</th>
<th>Senior (N=119)</th>
<th>All (N=624)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Preparation/Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper-higher</td>
<td>159</td>
<td>80.3</td>
<td>149</td>
<td>92.5</td>
<td>128</td>
</tr>
<tr>
<td>Paper-online-equal</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Online-higher</td>
<td>38</td>
<td>19.2</td>
<td>12</td>
<td>7.5</td>
<td>17</td>
</tr>
<tr>
<td>Material/Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper-higher</td>
<td>167</td>
<td>84.3</td>
<td>148</td>
<td>91.9</td>
<td>143</td>
</tr>
<tr>
<td>Paper-online-equal</td>
<td>1</td>
<td>.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Online-higher</td>
<td>30</td>
<td>15.2</td>
<td>13</td>
<td>8.1</td>
<td>3</td>
</tr>
<tr>
<td>Method/Skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper-higher</td>
<td>159</td>
<td>80.3</td>
<td>148</td>
<td>91.9</td>
<td>137</td>
</tr>
<tr>
<td>Paper-online-equal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Online-higher</td>
<td>39</td>
<td>19.7</td>
<td>12</td>
<td>7.5</td>
<td>9</td>
</tr>
<tr>
<td>Assignment/Examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper-higher</td>
<td>168</td>
<td>84.8</td>
<td>150</td>
<td>93.2</td>
<td>134</td>
</tr>
<tr>
<td>Paper-online-equal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Online-higher</td>
<td>30</td>
<td>15.2</td>
<td>11</td>
<td>6.8</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper-higher</td>
<td>170</td>
<td>85.9</td>
<td>152</td>
<td>94.4</td>
<td>140</td>
</tr>
<tr>
<td>Paper-online-equal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Online-higher</td>
<td>28</td>
<td>14.1</td>
<td>9</td>
<td>5.6</td>
<td>6</td>
</tr>
</tbody>
</table>

Conclusions and Recommendations

The purpose of this study has been to ascertain whether there is any significant difference in average ratings across survey administration methods. This study confirms that either paper or online method can be a valid and reliable way of student ratings of faculty teaching effectiveness. However, the mean scores of paper evaluation are significantly higher than those of online evaluation on each evaluation item, evaluation factor and the total ratings score. It seems that the classes give their instructors “less teaching effective scores” when they do student ratings with an online mode. A possible explanation may be that the online survey provides a safer and more candid response environment. Therefore, students who used the online method give fewer socially desirable
responses.

The results of this study suggest that online student ratings of instruction can be successfully administered at colleges where the student body is fairly computer literate and familiar with accessing the campus computer network. In addition, the study provided strong evidence that the validity and the reliability of student ratings do not vary according to an online or paper survey is used. This is consistent with the results of previous study (e.g. Layne, Decristoforo, and McGinty, 1999).

They are many survey methods for institutions to implement in order to attain student ratings of instruction. If institutions continue to believe in the importance of student voice in evaluating the faculty, it should be necessary to pay attention to the influence of survey methods on the rating result scores. Which scores are the real representation of faculty teaching effectiveness, paper scores or online scores? The educational institutions need to reflect deeply before they make a decision about their faculty members.

If the paper survey method is to be continuously implemented in most colleges as the mode of data collection for student ratings of faculty, the perception that it offers less anonymity must be countered. Students still believe that their instructors would find out what they had written about him/her on the paper evaluation form. Therefore, there is clearly an educational process that needs to take place in order to convince students that paper evaluations are legitimately anonymous. An endorsement by the student government association and other reputable agencies could help to allay fears about the paper survey system. In addition, an institutional guarantee of confidentiality might be necessary, similar to the guarantee that students’ records are confidential. Another way to avoid student giving social desirable ratings is to ask instructors to leave from the classroom while students are evaluating their instructors. Allowing students to do evaluation off campus is another way to administer the paper evaluation without involvement of faculty.

Though the response rate in the online group in this study was high, incentives may have to be offered to encourage students to respond when the online method is used in a non-forced implementation. A number of possible incentives could be tested experimentally to determine their effectiveness. For example, student who complete the online survey could be assigned earlier registration times for the next semester based on the number or proportion of courses that they evaluated during the current semester. Additionally, immediate access to course evaluation information could be made available to those students participating in the evaluation process. Higher response rates may also be achieved by allowing students to complete surveys somewhat earlier during the semester, when students are less pressured and the online resources are less likely to be overloaded. Student rating research has shown that rating results can be relatively stable from the midterm point to the end of the term (Costin, 1968; Feldman, 1979).

Overall, although the results are informative, they should be taken as a preliminary investigation. The generality of the findings should be strengthened with replications over different universities and other disciplines.
Acknowledgment

The author would like to thank the five AERA (American Educational Research Association) anonymous reviewers for their comments and National Hualien Teachers College in Taiwan for their willingness to participate in this research. This research was supported by research grants (NSC-90-2413-H-026-011) from the National Science Council.
References


Llewellyn, D. C. (2002). *Online reporting of student course survey results-methods, benefits and concerns*.


「學生評鑑教師教學」之結果：紙筆與網路調查的比較

張德勝
國立花蓮師範學院國民教育研究所

摘 要

本研究主要是比較紙筆調查與網路調查對於「學生評鑑教師教學」結果的差異性。研究對象為九十學年度第一學期國立花蓮師範學院大學部所開設的 624 門課，包含一年級 198 (31.73%) 門、二年級 161 (25.80%)、三年級 146 (23.40%) 及四年級 119 (19.07%) 門。研究工具為「國立花蓮師範學院教學意見反映調查表」共包含四個層面：「準備與計畫」、「教材與內容」、「教法與技巧」、「作業與評鑑」。

研究結果顯示在學生評鑑教師教學所有四個層面及總分，紙筆調查所得的結果都顯著的高於網路調查的結果。在所有的班級中，有 573 (91.8%) 門課，紙筆調查的平均分數高於網路調查的結果。相對的，只有 51 (8.2%) 門課，網路調查的結果高於紙筆調查的結果。研究結果顯示紙筆調查的方式比網路調查的方式，更容易讓學生給任課教師較高的評鑑分數。

關鍵字：學生評鑑教師教學、紙筆調查、網路調查