The practice of using student ratings to evaluate college teaching and studying factors which may affect the responses dates back to the early 1900s and the pioneering work of Remmers (1927, 1928, 1930) and his colleagues (Brandenburg & Remmers, 1927; Remmers & Brandenburg, 1927; Remmers, Martin, & Elliot, 1949). The body of knowledge related to traditional pencil-and-paper student evaluation of teaching (SET) ratings is broad and summaries of it have appeared over the years. For example, Centra (1993) reviewed what was known using four broad clusters of writing, including:

1. 1927 to 1960 when the work of Remmers, “The Father of Student Evaluation Research” and his colleagues at Purdue University was dominant;

2. 1960s when the use of student evaluations was almost entirely voluntary;
A Comparison of Web-Based and Paper-Based Course Evaluations

Valorie McAlpin, Mike Algozzine, Lee Norris, Richard Hartshorne, Richard Lambert, & Bob Algozzine

50 Fall 2014 (Volume 10 Issue 2)
Journal of Academic Administration in Higher Education

METHOD

Participants and Setting

We conducted our study at a large public urban university enrolling more than 25,000 students in the southeastern region of the United States. Each of the institution’s seven colleges (Architecture, Arts & Sciences, Business, Engineering, Health & Human Services) participated. Our research design sought participation from eight deans for each college, including two small (< 30) introductory undergraduate sections, two upper-level undergraduate sections, two graduate sections (> 10), and two graduate sections (> 10). Deans for each college presented the opportunity to take part in the pilot study to all eligible faculty in their college and participation was voluntary. From this, prospective participants from sections that met specific criteria (stratified courses) were selected and provided with a description of the project and the opportunity to participate. If any of the selected participants chose not to be included, additional participants were randomly selected from the list of volunteers. Section sizes below 10 were not included as they were considered exceptional and potentially different from other classes. As a result of logistical issues, one college had only seven courses and another college had only one course participating in a final sampling plan that included 48 course sections with 774 students randomly assigned to complete the course evaluations on-line and 779 randomly assigned to complete the course evaluations in-class. This blocking (i.e., assigning students to groups within sections of courses) controlled for instructor effects and was an important strength of our design.

We received usable evaluations (n = 1198, overall response rate of 77%) from courses taught by 41 instructors in 25 departments representing the following colleges: Architecture (16.2%), Arts & Sciences (22.7%), Business (6.9%), Computing & Informatics (13.9%), Education (14.4%), Engineering (13.9%), and Health & Human Services (14.9%). Of the usable evaluations, seven hundred and thirty-four (63.1%) of the evaluations were completed using the online administration and 451 (36.9%) were completed using the online administration. The distribution of responses across colleges and type of administration was not statistically significantly different, X²(6) = 4.55, p > .05.

Procedure

In-class course evaluations were conducted using instruments distributed and completed during class time in the traditional framework for campus-based courses (i.e., during a regularly scheduled meeting time in the session). Peers selected for the on-line evaluation participated in an electronic administration between a two-week window near the end of the semester. The greatest challenge in converting an on-line course evaluation system is the decline in student response rates that institutions often experience during the first year of transition; however, with a centrally-supported, controlled environment in which to administer course evaluations, student response rates generally remain in year two of the implementation (cf. Anderson, Cain, & Bird, 2005; Norris & Conn, 2005; Ravenscroft & Enyeart, 2009). Several additional factors are also issues requiring attention emerged in our study. To encourage participation, students in the on-line course evaluation group received up to six e-mail reminders, each containing a link to the evaluation instrument. Once students completed the survey, they did not receive additional reminder e-mails.

Instrumentation. Prior to implementing the study, we obtained current copies of course evaluation instruments from each participating college and department. These were then converted to electronic formats for the online evaluation group via a third-party vendor (Campus Lab). While there were a few university-required core evaluation items (e.g., Overall, this instructor was effective...), there was no common university-adopted instrument and the number (i.e., 7-27) and content of items varied across the participating departments and colleges; however, for this study, no modifications were made to the items or instruments submitted to the research team.

To reconcile data for subsequent analyses, two members of the research team independently identified common items representative of the following domains across the different evaluation instruments: Course purpose, positive learning environment, various instructional methods, use of instructional time, material relevance, lecture effectiveness, instructional effectiveness, instructor preparedness, instructor availability, grading fairness, grading usefulness, and overall satisfaction. For example, the “course purpose” item of the course purpose administration, included a course being offered at a particular time of day or night, or online as an option for that particular course. Half of the students in a section of the course being offered at a particular time of day or night or online at the same time as another course. The other half completed the traditional in-class course evaluations. By doing this, we controlled for “teacher effects” in that every instructor was rated by students in both the on-line and in-class courses. A total of 775 randomly assigned to complete the course evaluations...
A Comparison of Web-Based and Paper-Based Course Evaluations

RESULTS

Response Rates

A total of 1,549 students were randomly assigned within the participating courses to complete their course evaluations in-class using the paper-based procedures or to complete their course evaluations through the on-line system. \( n_{\text{in-class}} = 775, n_{\text{on-line}} = 774 \). A total of 1,171 students (\( n_{\text{in-class}} = 714, n_{\text{on-line}} = 457 \)) provided sufficient information to be included in the analysis. At least five students responded in 39 different courses; however, one course was dropped from the analysis as only two students responded and a small number of students were dropped from the analyses (\( n = 25 \)) because of incomplete data. The response rate was very high for the in-class condition (92.13%) and lower for the on-line condition (59.84%).

A number of faculty participants cited confusion with the selection of the on-line participants (e.g., students were not sure if they received the e-mail). This may have had an effect on the response rates in the study, as faculty noted the possibility of confused students accidentally completing the in-class course evaluations, even though they were in the group designated to complete the on-line student course evaluations. Students were likewise confused by receiving email from Campus Labs to notify or remind them to complete the web-based evaluation. Since they were not familiar with Campus Labs, many of them may have treated the reminders as spam and likely never completed the evaluation. This could have had a significant impact on response rate, since the emails did not come directly from the university.

Ratings

The level-one, within-course variance, models, included the scale scores constructed from the course evaluation items as the dependent variables. A separate model was conducted for each outcome measure. Treatment group membership was entered as an uncentered predictor variable in the level-one models. The level-two models, the between-course models, were unconditional models with no predictor variables. Completely unconditional models were calculated as the first step in the analysis and 79.1% of the variance in course evaluation ratings was found to be within courses, while 20.1% of the variance in the ratings was between courses. In general, average ratings across group and area of rating were above 4 (on the 5-point scale), reflecting positive evaluations. There was a small, statistically significant difference, \( t = 2.44, p < .05 \) between the groups on overall satisfaction; ratings for the in-class group (\( M = 4.43, SD = 0.64 \)) were slightly higher than those for the on-line group (\( M = 4.40, SD = 0.66 \)); however, when expressed as a standardized mean difference effect size based on the pooled within course standard deviation estimates from the HLM models, the practical significance of the difference was small (\( d = 0.16 \) and \( 0.00 \) was included in the 95% confidence interval. Students in both conditions were, on average, positive about the course experience. All scale score means, across both groups, were not lower than 4 on the 5-point scale. As shown in Table 1, a similar pattern of small, statistically significant differences was found for 9 of the 12 scale scores. For the remaining three scale scores, there was not a statistically significant difference between the groups. In general, the differences between ratings obtained using in-class and on-line evaluations were small (Range = .08 to .99 on 5-point scale), and, for none of the scale scores were the between group differences exceeding an effect size of approximately .20. We also compared the distribution of very low and very high ratings across our groups. As illustrated in Table 2, “strong” opinions (i.e., ratings of 1 or 5), were similarly distributed across in-class and on-line evaluations. Coupling these findings with the possibility that the statistically significant differences were due in part to the large sample sizes in our analyses, we judged the practical and observed value of all of the group differences to be small (see Figure 1).

Costs

We reasoned that on-line course evaluations would generate substantial savings to the institution for materials and staff time (see Figure 2). Conservative estimates indicate that 80 hours of departmental staff time from each of 80 staff members is required to complete paper-based course evaluations with an annual cost of $224,000 for

### Table 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Area of Rating</th>
<th>Grading Fairness</th>
<th>Grading Usefulness</th>
<th>Course Purpose</th>
<th>Use of Instructional Time</th>
<th>Instructor Availability</th>
<th>Overall Satisfaction</th>
<th>Material Relevance</th>
<th>Learning Effectiveness</th>
<th>Varied Instructional Methods</th>
<th>Positive Learning Environment</th>
<th>Instructional Effectiveness</th>
<th>Instructor Preparedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-Class</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>On-Line</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( 95\% \text{ CI} \)

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>U</th>
<th>L</th>
<th>U</th>
</tr>
</thead>
</table>

1. \( ES^2 = d^2 = (M_{\text{In-Class}} - M_{\text{On-line}}) / SD_{\text{Pooled}} \), where \( d \) reflects small practical difference (cf. Cohen, 1988)

2. \( p < .05 \)

### Table 2

<table>
<thead>
<tr>
<th>Course Purpose</th>
<th>Description Qty</th>
<th>Cost/Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-Class</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>On-Line</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Instructional Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Availability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Satisfaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material Relevance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Effectiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varied Instructional Methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Learning Environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Effectiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Preparedness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Costs of In-Class vs. On-Line Student Course Evaluations**

**Figure 1**

**Figure 2**

**Figure 1: In-Class vs. On-Line Student Course Evaluation Ratings Comparison**

**Figure 2: Cost/Savings of In-Class vs. On-Line Student Course Evaluations**
A Comparison of Web-Based and Paper-Based Course Evaluations

Valorie McAlpin, Mike Algozzine, Lee Norris, Richard Hartshorne, Richard Lambert, & Bob Algozzine

Personnel cost projections derive from estimates of personalized paper forms; $5,000 in licensing costs for the existing web-based evaluation system currently used for distance education courses (i.e., this cost would be removed if the entire campus went to web-based student course evaluations); and $6,246 in OPSCAN personnel costs (total annual cost is $250,246). The cost of licensing web-based course evaluation software for the entire university is $24,500 annually. Coupled with the survey administration and management costs of $56,500, we estimated that the university could realize a cost savings of $169,246, or a 68% savings in the operating costs of the student course evaluation process (i.e., a five-year savings of more than three-quarters of a million dollars).

Additionally, decentralized student evaluation systems lack uniform administrative support, which makes university-wide data comparisons of faculty teaching difficult and unwieldy when provisions for administrative oversight, support, and coordination have not been considered.

The on-line course evaluation method has several benefits to faculty, students, and the institution, including:

- Shorter turnaround time to deliver feedback to faculty, department chairs, and deans.
- Increased ability to perform statistical analyses with course evaluation data.
- Improved ability to perform longitudinal comparisons of institutional and individual results.
- Improved ability for individual faculty to evaluate results across all their assigned courses.
- More substantive feedback from students on open-ended questions.
- Increased efficiency from less manual manipulation required by administrative staff.
- Better data, since errors are less likely and open-ended responses are generally more complete.
- Open-and continuous-access for students rather than attendance-based opportunity restricted to a single day in class.
- Substantial savings to the institution for materials and staff time, including reduced printing, distribution, collection, and storage costs.

Additionally, while a detailed quantitative and qualitative analysis of the open-ended responses is ongoing, a cursory review of these responses indicated that there was a significant increase in the quantity of open-ended responses on the online student course evaluations. This was even more significant, as a number of the participating departments omitted the open-ended responses from their pencil-and-paper evaluation instruments. This preliminary post hoc finding aligns with previous reports that cite additional resources of staff time, and delays feedback to course instructors.

In a recent study, Young and McCaslin (2013) compared course evaluation data across the institution can create considerable variability in personnel cost estimates.

Personnel cost projections derive from estimates by departmental assistants involved in study. Variation across the institution can create considerable variability in personnel cost estimates.

The magnitude of the differences aside, variations in ratings may reflect different contextual opportunities provided by in-class and web-based evaluation administrations. For example, students may think more negatively given more time and distance from the instructor when evaluating a course outside the classroom. Although expectations are that instructors are not present during in-class course evaluation administrations, the perception of more anonymity online may also have been a source of variation across scores in our study. Again, the obtained differences between ratings on in-class and in-class assessments were small; however, additional randomized controlled trials are warranted to support future decisions and policy related to this important higher education practice.

Because the domains selected typically resulted in the favorable ratings noted above, these small differences across methods should not surprise administrators or faculty. More important from a policy perspective, the in-class course evaluation method has several limitations, including:

- Allocating materials escalates institutional costs needed for paper, printing, distribution, collection, scoring, reporting, and storage.
- Transcribing comments creates opportunities for subjective interpretations based on the quality of the handwriting, requires additional resources of staff time, and delays feedback to course instructors.
- Administering evaluations in the classroom limits the amount of time students are able to dedicate to the evaluations, requires devoting a portion of class time to completing evaluations, and poses limitations on the effectiveness of the evaluations (i.e., students complain of being unable to contribute thoughtful comments in a short timeframe).

Additionally, decentralized student evaluation systems lack uniform administrative support, which makes university-wide data comparisons of faculty teaching difficult and unwieldy when provisions for administrative oversight, support, and coordination have not been considered. The on-line course evaluation method has several benefits to faculty, students, and the institution, including:

- Shorter turnaround time to deliver feedback to faculty, department chairs, and deans.
- Increased ability to perform statistical analyses with course evaluation data.

REFERENCES

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Grettas, J., Howley, L., ... & Spooner, F. (2004). Student evaluation research would do well to a formal study of large number of classes within the university... (p. 16). (e.g., Liberal Arts and Science, Engineering, Education and instruction in more than 30 courses), there were small, statistically significant differences that slightly favored the in-class student course evaluations; however, given the large sample size and the consistently low effect sizes, there was low practical significance in the difference in the ratings. The magnitude of the differences aside, variations in ratings may be due to unique and different contextual opportunities provided by in-class and web-based evaluation administrations. For example, students may think more negatively given more time and distance from the instructor when evaluating a course outside the classroom. Although expectations are that instructors are not present during in-class course evaluation administrations, the perception of more anonymity online may also have been a source of variation across scores in our study. Again, the obtained differences between ratings on in-class and in-class assessments were small; however, additional randomized controlled trials are warranted to support future decision making and policy related to this important higher education practice.

Because the domains selected typically resulted in the favorable ratings noted above, these small differences across methods should not surprise administrators or faculty. More important from a policy perspective, the in-class course evaluation method has several limitations, including:

- Allocating materials escalates institutional costs needed for paper, printing, distribution, collection, scoring, reporting, and storage.
- Transcribing comments creates opportunities for subjective interpretations based on the quality of the handwriting, requires additional resources of staff time, and delays feedback to course instructors.
- Administering evaluations in the classroom limits the amount of time students are able to dedicate to the evaluations, requires devoting a portion of class time to completing evaluations, and poses limitations on the effectiveness of the evaluations (i.e., students complain of being unable to contribute thoughtful comments in a short timeframe).

Additionally, decentralized student evaluation systems lack uniform administrative support, which makes university-wide data comparisons of faculty teaching difficult and unwieldy when provisions for administrative oversight, support, and coordination have not been considered. The on-line course evaluation method has several benefits to faculty, students, and the institution, including:

- Shorter turnaround time to deliver feedback to faculty, department chairs, and deans.
- Increased ability to perform statistical analyses with course evaluation data.

Improved ability to perform longitudinal comparisons of institutional and individual results.

Improved ability for individual faculty to evaluate results across all their assigned courses.

More substantive feedback from students on open-ended questions.

Increased efficiency from less manual manipulation required by administrative staff.

Better data, since errors are less likely and open-ended responses are generally more complete.

Open-and continuous-access for students rather than attendance-based opportunity restricted to a single day in class.

Substantial savings to the institution for materials and staff time, including reduced printing, distribution, collection, and storage costs.

Additionally, while a detailed quantitative and qualitative analysis of the open-ended responses is ongoing, a cursory review of these responses indicated that there was a significant increase in the quantity of open-ended responses on the online student course evaluations. This was even more significant, as a number of the participating departments omitted the open-ended responses from their pencil-and-paper evaluation instruments. This preliminary post hoc finding aligns with previous reports that cite additional resources of staff time, and delays feedback to course instructors.

In a recent study, Young and McCaslin (2013) compared course evaluation data...
A Comparison of Web-Based and Paper-Based Course Evaluations

Valerie McAlpin, Mike Algozzine, Lee Norris, Richard Hartshorne, Richard Lambert, & Bob Algozzine

Fall 2014 (Volume 10 Issue 2)
egies for Increasing Student Participation Rates. pdf


