A Comparison of Web-Based and Paper-Based Course Evaluations

Valorie McAlpin
Center for Teaching and Learning
University of North Carolina at Charlotte

Mike Algozzine
Center for Teaching and Learning
University of North Carolina at Charlotte

Lee Norris
Center for Teaching and Learning
University of North Carolina at Charlotte

Richard Hartshorne
Department of Educational & Human Sciences
University of Central Florida

Richard Lambert
Center for Educational Measurement and Evaluation
University of North Carolina at Charlotte

Bob Algozzine
Center for Educational Measurement and Evaluation
University of North Carolina at Charlotte

The authors gratefully acknowledge the assistance of J. Garvey Pyke and Tom Cook in reviewing early project-related reports that provided a base for this article and in editing drafts of our manuscript. The opinions expressed herein are those of the authors, and no official endorsement should be inferred.

ABSTRACT

Student evaluations of college teaching have been endorsed and criticized for as long as they have been used as part of important decision-making practices in higher education. With the growth of distance education, the need for alternative approaches for these assessments has increased. We were interested in the extent to which outcomes were comparable across in-class and on-line course evaluations. We conducted a randomized controlled trial across 7 colleges, 25 departments, and 41 instructors at a large urban research university in the southeastern part of the United States. The distribution of ratings across demographic and comparison groups was similar. Response rates were lower for students participating online; however, none of the scale score differences between groups exceeded an effect size .21 and the estimated benefits were large. We discuss the advantages and disadvantages of alternative approaches for evaluating instruction in the context of past, current, and future research and practice.

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;

3. 1970s when the focus was on demonstrating the technical adequacy and usefulness of ratings; and,

4. 1980s to the then present day when the research provided continued clarification and amplification of prior findings with syntheses of extant studies as well as new investigations. (p. 49)

Using several articles published in the *American Psychologist* as a base, McKeachie (1997) summarized opinions and evidence related to the number of dimensions of SET ratings that should be used in personnel decisions, the validity of the ratings relative to teaching effectiveness, and the potential for controlling biases if they are evident in the ratings. More recently, Sproule (2000) reviewed methodological concerns related to student evaluations of teaching and Algozzine et al. (2004) summarized what was known about evaluating "...the effectiveness of instruction in postsecondary education and proposed areas for improvements, as well as considerations for future research" (p. 1). The knowledge base here is presented positively by some (cf. d'Apollonia &Abrami, 1997; Gillmore, 1984; Greenwald & Gilmore, 1997; Marsh, 1987; Marsh & Roche, 1997; McKeachie, 1997; Ramsden, 1991; Ruskai, 1996; Seldin, 1989, 1998; Shingles, 1977; Trujillo, 1986; Wachtel, 1998) and equivocally or negatively by others (Algozzine, Beattie, Bray, Flowers, Gretes, Mohanty, & Spooner, 2010; Centra, 1979; Damron, 1995; Haskell, 1997a, b, c, d; Mohanty, Gretes, Flowers, Algozzine, & Spooner, 2005, 2006; Young & McCaslin, 2013). Regardless of arguable strengths or weaknesses, based on longevity alone, student ratings of instruction remain "... an unavoidable reality of higher education and the messages communicated...in them often play a role in merit, promotion and tenure decisions" (Vennette, Sellnow, & McIntyre, 2010, p. 102). The constancy and power of this practice is driving new interest in the methods of delivery used to collect course evaluation ratings in both distance education and traditional campus-based courses (cf. Anderson, Brown, & Spaeth, 2006; Anderson, Cain, & Bird, 2005; Avery, Bryant, Mathios, Kang, & Bell, 2006; Cohen, Carbone, Beffa-Negrini, 2001; Crews & Curtis, 2011; Dommeyer, Baum, & Hanna, 2002; Dommeyer, Baum, Hanna, & Chapman, 2004; Donovan, Mader, & Shinsky, 2006; Harrington & Reasons, 2005; Hmielseski & Champagne, 2000; Johnson, 2003; Kanagaretnam, Mathieu, & Thevaranjan, 2003; Kasiar, Schroeder, & Holstaad, 2001; Kuhtman, 2004; Layne, DeCristoforo, & McGinty, 1999; Morrison, 2011; Sorenson & Johnson, 2003; Stewart, Waight, Marcella, Norwood, & Ezell, 2004; Venette, Sellnow, & McIntyre, 2010).

Granello and Wheaton (2004) point out that web-based data collection procedures offer a number of positive features such as "...reduced response time, lower cost, ease of data entry, flexibility of and control over format, advances in technology, recipient acceptance of the format, and the ability to obtain additional response-set information" (p. 388). In the developing world of online technologies, it is no surprise that Internet-based surveys are being considered on campuses across the country as alternatives to traditional pencil-and-paper methods when conducting end-of-course student evaluations of instruction; but, again, the knowledge base is equivocal. For example, while convenience, completeness, efficiency, cost-effectiveness, and student preference are among positive features, concerns related to technology, higher percentage of negative responses, and lower response rates have dampened the ease and speed with which online assessments have been deemed acceptable to faculty and other decision makers (Anderson, Cain, & Bird, 2005; Carini, Hayek, Kuh, Kennedy, Ouimet, 2003; Dommeyer, 2006; Dommeyer, Baum, Chapman, & Hanna, 2002; Donovan, Mader, & Shinsky, 2005; Paolo, Bonaminio, Gibson, Partridge, & Kallail, 2000; Seok, DaCosta, Kinsell, & Tung, 2010; Sorenson & Johnson, 2003; Venette, Sellnow, & McIntyre, 2010; Watt, Simpson, McKillop, & Nunn, 2002; Winer & Sehgal, 2006).

To address challenges associated with the ongoing implementation of student evaluations of teaching, we explored the use of an online alternative in a campus-wide study. We were interested in the extent to which response rates, ratings, and costs were comparable across in-class and on-line administrations of course evaluations. We used existing structures and practices within our university to complete the study.

METHOD

Participants and Setting

We conducted our study at a large public urban research 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, Computing & Informatics, Education, Engineering, Health & Human Services) participated.

Our research design sought participation from eight course sections (i.e., group of students taking a course at a particular time of day or night) from each college, including two small (n < 30) introductory undergraduate sections, two large (n > =30) introductory undergraduate sections, two upper-level undergraduate sections (n > 10), and two graduate sections (n > 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 participate and another college had only one course participate resulting in a final sampling plan that included 48 course sections with 774 students randomly assigned to complete the course evaluations on-line and 775 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 (4.0%), 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 (61.3%) of the evaluations were completed using the traditional in-class method and 464 (38.7%) were completed using the online administration. The distribution of responses across colleges and type of administration was not statistically significantly different, $X^2(6) = 4.55$, $p > .05$.

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 session near the end of the semester). Peers selected for the on-line evaluation participated in an electronic administration during a two-week window near the end of the semester.

The greatest challenge in converting to 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 return in year two to the previous rates (cf. Anderson, Cain, & Bird, 2005; Norris & Conn, 2005; Ravenscroft & Enyeart, 2009). Several additional potential 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 Labs*). While there were a few university-required core evaluation items (e.g., Overall, I learned a lot in this course. 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, varied instructional methods, use of instructional time, material relevance, learning effectiveness, instructional effectiveness, instructor preparedness, instructor availability, grading fairness, grading usefulness, and overall satisfaction. For example, the "course purpose" item (i.e., The course has clearly stated objectives) was item 8 on the College of Architecture instrument, item 6 on the Business Administration Marketing Department instrument, and item 7 on the College of Education instrument. We then compared the overall satisfaction score and the 11 domain scores across web-based and paper-based groups.

Design and Data Analysis

The research design was a randomized controlled trial (RCT) of students assigned to in-class or on-line course evaluation administrations. Half of the students in a section of the a course being offered at a particular time of day or night piloted the on-line course evaluation and 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 group. Since students were nested within courses, rating comparisons between the two treatment conditions were completed by using multilevel modeling techniques (Bickel, 2007). In the cases where there were multiple sections for a given course, the sections were combined. An average of 30 students responded per course (minimum = 6, maximum = 109). Data analysis included comparisons of responses rates and ratings obtained using different methods and a prospective analysis of the cost-benefits of using online evaluations. We used

the .05 level of significance; and, calculated effect sizes adjusted for the clustering effects of the nested design (i.e., ES = group differences divided by the model-estimated pooled within group standard deviation from HLM analyses) and confidence intervals (CI) to document the statistical and practical levels of obtained differences (cf. Cohen, 1988; Peugh, 2010; Roberts & Monaco, 2006; Thompson, 2006).

We believe that the research design selected (i.e., randomly assigning participants within courses to each group rather than selecting entire courses to complete either the student course evaluation on-line or in-class) was more rigorous and provided us with more powerful results than reported in prior research. Another design concern was the lack of a common course evaluation instrument. In attempting to reconcile the data for analysis, it was obvious that the content of student course evaluations from each college varied a great deal and was designed to measure very different aspects of teaching and learning. Thus, we had to derive common themes reflective of 12 domains of interest rather than use responses to the same items for comparisons of ratings across methods. We do not believe that this greatly restricted our findings given the large number of individual responses that contributed to our comparisons.

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 process 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 ($n_{\text{In-class}} = 714$, $n_{\text{Online}} = 457$) 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 (59.04%).

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-mails). 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 ($d = .16$) 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* = -.07 to .09 on 5-point scale); and, for none of the scale scores were the between group differences exceeding an effect size of approximately .21. 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).

Table 1
Comparison of Student Evaluations across Administration Method

	Group						95% CI	
	In-Class		On-Line					
Area of Rating	M	SD	M	SD	*t*	*ES*[1]	LL	UL
Grading Fairness	4.26	0.94	4.33	0.93	1.16	.08	-.04	.19
Grading Usefulness	4.30	0.88	4.21	1.02	2.38[2]	.17	.05	.29
Course Purpose	4.34	0.84	4.27	0.95	2.21[2]	.15	.03	.26
Use of Instructional Time	4.35	0.89	4.30	0.93	2.85[2]	.21	.09	.33
Instructor Availability	4.37	0.82	4.35	0.89	1.65	.12	.01	.24
Overall Satisfaction	4.43	0.64	4.40	0.66	2.44[2]	.16	.04	.28
Material Relevance	4.46	0.76	4.37	0.89	2.75[2]	.17	.06	.29
Learning Effectiveness	4.46	0.83	4.37	0.90	2.73[2]	.17	.06	.29
Varied Instructional Methods	4.47	0.78	4.41	0.86	2.81[2]	.16	.04	.27
Positive Learning Environment	4.50	0.76	4.46	0.86	2.11[2]	.13	.01	.25
Instructional Effectiveness	4.51	0.82	4.46	0.84	2.35[2]	.16	.04	.28
Instructor Preparedness	4.55	0.66	4.49	0.76	1.58	.15	.03	.26

1 *ES* (*Effect Size*) = *d* = (MIn-Class – MOn-line)/ SDPooled, where .20 reflects small practical difference (cf. Cohen, 1988)
2 $p < .05$

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

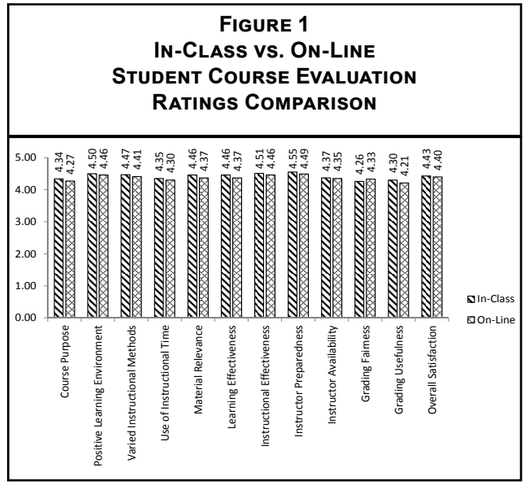

Figure 2
Cost/Savings of In-Class vs. On-Line Student Course Evaluations

Course Evaluation Cost Analysis

Description	Qty	Cost Per	Total
In-Class Cost			
Cost of Paper Forms [including overprint]	100,000	$ 0.15	$ 15,000.00
Software Licensing Distance Education On-line Course Evaluation	1	$5,000.00	$ 5,000.00
Departmental Staff Processing Time			
(80 staff members @ 80 hours each for processing written comments	6,400	$ 35.00	$ 224,000.00
Reduction (37%) in OPSCAN Availability	488	$ 12.80	$ 6,246.40
Annual In-Class Cost Estimate			$ 250,246.40
On-Line Cost/Savings			
On-Line Software			$ 24,500.00
Institutional Administration and Management			$ 56,500.00
(Paper)			$ (15,000.00)
(Software License)			$ (5,000.00)
(Staffing)			$ (224,000.00)
(OPSCAN)			$ (6,246.40)
Annual On-Line Cost/Savings Estimate			$ (169,246.40)
Savings Summary			
Annual In-Class Evaluation Costs			$ 250,246.40
Annual On-Line Cost/Savings Estimate			$ (169,246.40)
Percent Reduction in Costs			68%
Five-Year Savings Estimate			$ (846,232.00)

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 2
Percent of Low and High Ratings across Paper- and Web-Based Administrations

	Rating			
	Low		High	
Area of Rating	In-Class	On-Line	In-Class	On-Line
Overall Satisfaction	0.1%	0.2%	32.5%	25.9%
Instructor's Preparedness	0.3%	0.3%	62.8%	62.8%
Instructor's Availability	0.6%	1.4%	55.1%	56.0%
Positive Learning Environment	0.7%	1.9%	62.4%	62.0%
Materials Relevance	0.8%	1.1%	58.5%	56.9%
Grading Fairness	0.9%	1.9%	51.8%	54.8%
Varied Instructional Materials	1.0%	1.5%	60.7%	58.0%
Course Purpose	1.2%	2.1%	52.6%	52.3%
Instructional Effectiveness	1.3%	1.5%	65.6%	61.6%
Use of Instructional Time	1.3%	1.7%	56.1%	53.6%
Learning Effectiveness	1.4%	1.5%	61.3%	57.5%
Grading Usefulness	1.6%	4.0%	51.6%	49.8%

personnel[1]. Additional costs include $15,000 for customized 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 would 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).

DISCUSSION

In a recent study, Young and McCaslin (2013) compared student evaluations of faculty in a college of business administration using "traditional in-class" and "online" methods and found no "significant differences in mean scores...in the majority of cases" (p. 11). A "major limitation of this study was the use of only eight classes within one college..." and the researchers indicated that "[f]uture 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 created by on-line and in-class course 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 on-line 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 time as a key indicator of both quality and quantity of open-ended responses as well as with prior findings that transcription and other errors are less likely and open-ended responses are generally more detailed when completed using online evaluation methods (cf. Kasiar, Schroeder, & Holstaad, 2001; Layne, DeCristoforo, & McGinty, 1999; Ravelli, 2000; Venette, Sellnow, & McIntyre, 2010; Young & McCaslin, 2013).

While response rate differences for in-class and on-line administrations in our study may be a function of the experimental nature of work and may disappear when a single option is offered, achieving adequate response rates and identifying strategies to improve them is a consistently reported faculty concern (cf. Crews, 2011; Dommeyer, Baum, Chapman, & Hanna, 2002). Additional challenges and potential disadvantages include the need to obtain faculty buy-in, responding to faculty and student concerns for anonymity and privacy, and changing the culture of higher education to support on-line student evaluation of teaching (New Jersey Institute of Technology, 2008).

CONCLUSION

Our research was designed to examine commonly-reported concerns and other issues related to the implementation of on-line student course evaluations. We believe our work provides guidance for faculties interested in exploring the use of on-line student course evaluations as an alternative for in-class paper-pencil scan-sheet methods. More specifically, the foundations of information provided to faculty councils and other decision-making bodies for review, consideration, and consultation regarding future changes in student evaluation of teaching procedures should include sufficient evidence of similarities and differences in response rates between in-class and on-line evaluation formats; documentation of the extent to which ratings are comparable between in-class and on-line formats; analysis of similarities and differences in qualitative feedback to determine if evaluation delivery medium impacts results; and, support for the cost-efficiency of resource use between in-class and on-line formats.

REFERENCES

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., ... & Spooner, F. (2004). Student evalua-

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

tion of college teaching: A practice in search of principles. *College Teaching*, *52*, 134-141.

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Mohanty, G., & Spooner, F. (2010). Multi-method evaluation of college teaching. *Journal of Excellence in College Teaching*, *21*, 27-49.

Anderson, H. M., Cain, J., & Bird, E. (2005). Online course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education, 69, 34-43.*

Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. Innovate 2(6). Retrieved from http://www.innovateonline.info/pdf/vol2_issue6/Online_Student_Evaluations_and_Response_Rates_Reconsidered.pdf

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Electronic Education, 37, 21-38.*

Bickel, R. (2007). *Multilevel analysis for applied research.* New York: Guilford.

Brandenburg, G. C., & Remmers, H. H. (1927). A rating scale for instructors. *Educational Administration and Supervision, 13*, 399-406.

Carini, R., Hayek, J., Kuh, G., Kennedy, J., & Ouimet, J. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education, 44, 1-19.*

Centra, J.A. (1979). *Determining faculty effectiveness: Assessing teaching, research, and service for personnel decisions and improvement.* San Francisco: Jossey-Bass.

Centra, J.A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (*2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, N. L., Carbone, E. T., & Beffa-Negrini, P. A. (2011). The design, implementation, and evaluation of online credit nutrition courses: A systematic review. *Journal of Nutrition Education & Behavior*, *43*(2), 76-86. doi:10.1016/j.jneb.2010.04.001

Crews, T. B., & Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education,* 36(7), 865-878. doi:10.1080/02602938.2010.493970

Damron, J. C. (1995). The three faces of teaching evaluation. Unpublished manuscript. New Westminster, British Columbia: Douglas College.

d'Apollonia, S., & Abrami, P. (1997). Navigating student ratings of instruction. *American Psychologist*, *52, 1198-1208.*

Dommeyer, C. J. (2006). The effect of evaluation location on peer evaluations. *Journal of Education for Business, 82, 21-26.*

Dommeyer, C. J., Baum, P., & Hanna, R. W. (2002). College students' attitudes toward methods of collecting teaching evaluations: In-class versus on-line. *Journal of Education for Business*, 78, 11-15.

Dommeyer, C. J., Baum, P., Chapman, K., & Hanna, R. (2002). Attitudes of business faculty toward two methods of collecting teacher evaluations: Paper vs. online. *Assessment and Evaluation in Higher Education*, 27, 455-462.

Dommeyer, C. J., Baum, P., Hanna, R., & Chapman, K. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education,* 29, 611-623.

Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive student feedback: Online vs. Traditional course evaluations. *Journal of Interactive Online Learning*, 5, 283-296.

Gillmore, G. (1984). Student ratings as a factor in faculty employment decisions and periodic review. *Journal of College and University Law,* 10, 557- 576.

Granello, D., & Wheaton, J. E. (2004). Online Data collection: Strategies for research. *Journal of Counseling & Development,* 82, 387-393. doi:10.1002/j.1556-6678.2004.tb00325.x

Greenwald, A., & Gilmore, G. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.

Harrington, C. F., & Reasons, S. G. (2005). Online student evaluation of teaching for distance education: A perfect match? *The Journal of Educators Online, 2(1),* 1-12. Retrieved from http://www.thejeo.com/ReasonsFinal.pdf

Haskell, R.E. (1997a). Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives,* 5(6). Retrieved from http://epaa.asu.edu/epaa/v5n6.html

Haskell, R.E. (1997b). Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part II) Views from court. *Education Policy Analysis Archives* 5(6). Retrieved from http://epaa.asu.edu/epaa/v5n17.html

Haskell, R.E. (1997c). Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part III) Analysis and implications of views from the court in relation to accuracy and psychometric validity. *Education Policy Analysis Archives,* 5(6).Retrieved from http://epaa.asu.edu/epaa/v5n18.html

Haskell, R.E. (1997d). Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part IV) Analysis and implications of views from the court in relation to academic freedom, standards, and quality of instruction. *Education Policy Analysis Archives,* 5(6). Retrieved from http://epaa.asu.edu/epaa/v5n21.html

Hmielseski, K., & Champagne, M. V. (2000, September/October). Plugging in to course evaluation. *The Technology Source*. Retrieved from http://technologysource.org/article/plugging_in_to_course_evaluation/

Hobson, S. M., & Talbot, D. N. (2001). Understanding student evaluations: What all faculty should know. *College Teaching,* 49, 26-31.

Johnson, T. D. (2003). Online student ratings: Will students respond? New directions for teaching and learning: Online student ratings of instruction, 96, 49-59.

Kanagaretnam, K., Mathieu, R., & Thevaranjan, A. (2003). An economic analysis of the use of student evaluations: Implications for universities. *Managerial and Decision Economics*, 24, 1-13.

Kasiar, J.B., Schroeder, S. L., & Holstaad, S. G. (2001). Comparison of traditional and web-based course evaluation processes in a required, team-taught pharmacotherapy course. *American Journal Pharmaceutical Education*, 66, 268-270.

Kuhtman, M. (2004) Review of online student ratings of instruction. *College and University Journal*. 80, 64-67.

Layne, B.H., DeCristoforo, J.R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40, 221-232.

Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research,* 11, 253-388.

Marsh, H., & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: The central issues of validity, bias, and utility. *American Psychologist, 52*, 1187-1197.

McKeachie, W. (1997). Student ratings: The validity of use. *American Psychologist*, *52*, 1218–1225.

Mohanty, G., Gretes, J., Flowers, C., Algozzine, B., & Spooner, F. (2005). Multi-method evaluation of instruction in engineering classes. *Journal of Personnel Evaluation in Education*, 18, 139-151. [Also published online May 26, 2006]

Mohanty, G., Gretes, J., Flowers, C., Algozzine, B., & Spooner, F. (2006, May 26). Multi-method evaluation of instruction in engineering classes. *Journal of Personnel Evaluation in Education*. Available from http://dx.doi.org/10.1007/s11092-006-9006-3 [Also published 2005, 18, 139-151.

Morrison, R. (2011). A comparison of online versus traditional student end-of-course critiques in resident courses. *Assessment & Evaluation in Higher Education,* 36, 627-641.

New Jersey Institute of Technology. (2008). Online course evaluation. Retrieved from http://www.njit.edu/about/pdf/online-student-evaluation-supporting-material-2008-10-13.pdf

Norris, J., & Conn, C. (2005). Investigating strategies for increasing student response rates to online-delivered course evaluation. *Quarterly Review of Distance Education*, 6, 13-29.

Paolo, A. M., Bonaminio, G. A., Gibson, C., Partridge, T., & Kallail, K. (2000). Response rate comparison of e-mail and mail-distributed student evaluations. *Teaching & Learning in Medicine*, 12(2), 81-84.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85-112.

Ramsden, P. (1991) Performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16, 129-150.

Ravelli, B. (2000, June). Anonymous online teaching assessments: Preliminary findings. Paper presented at the Annual National Conference of the American Association for Higher Education, Charlotte, NC.

Ravenscroft, M. & Enyeart C. (2009, May 8). Online student course evaluations: Strategies for increasing student participation rates (Custom Research Brief). Washington, D.C.: The Advisory Board Company. Retrieved from http://tcuespot.wikispaces.com/file/view/Online+Student+Course+Evaluations+-+Strat

egies+for+Increasing+Student+Participation+Rates.pdf

Remmers, H. H. (1927). The Purdue Rating Scale for Instructors. *Educational Administration and Supervision*, 6, 399-406.

Remmers, H. H. (1928). The relationship between students' marks and students' attitudes toward instructors. *School and Society*, 28, 759-760.

Remmers, H. H. (1930). To what extent do grades influence student ratings of instructors? *Journal of Educational Psychology*, 21, 314-316.

Remmers, H. H., & Brandenburg, G. C. (1927) Experimental data on the Purdue ratings scale for instructors. *Educational Administration and Supervision*, 13, 519-527.

Remmers, H. H., Martin, F. D., & Elliot, D. N. (1949). Are student ratings of their instructors related to their grades? *Purdue University Studies in Higher Education*, 44, 17-26.

Roberts, J. K. & Monaco, J. P. (April, 2006). Effect size measures for the two level linear multilevel model. Paper presented at the Annual meeting of the American

Educational Research Association, San Francisco, CA.

Ruskai, M.B. (1996). Evaluating student evaluations. *Notices of the American Mathematical Society*, 44, 308.

Seldin, P. (1989). How colleges evaluate professors. 1988 vs. 1983. *AAHE Bulletin,* 41(7), 3-7.

Seldin P. (1998) Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23, 191-212.

Seok, S., DaCosta, B., Kinsell, C., & Tung, C. K. (2010). Comparison of instructors' and students' perceptions of the effectiveness of online courses. *Quarterly Review of Distance Education,* 11, 25-36.

Shingles, R. D. (1977). Faculty ratings: Procedures for interpreting student evaluations. *American Educational Research Journal,* 14, 459-470.

Sorenson, D.L. & Johnson, D. (2003). *Online student ratings of instructions*. San Francisco: Jossey Bass.

Sproule, R. 2000. Student evaluation of teaching: A methodological critique of evaluation practices. *Education Policy Analysis Archives*, 8(50). Retrieved from http://epaa.asu.edu/epaa/v8n50.html

Stewart, B. L., Waight, C. L., Marcella, W., Norwood, M. M., & Ezell, S. D. (2004). Formative and summative evaluation of online courses. *Quarterly Review of Distance Education,* 5, 101-109.

Thompson, B. (2006). *Fundamentals of behavioral statistics.* New York: The Guilford Press.

Trujillo, C. M. (1986). A comparative examination of classroom interactions between professors and minority and non-minority students. *American Educational Research Journal,* 23, 629-642.

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, 35, 101-115.

Wachtel, H.K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education,* 23, 191-211.

Watt, S., C. Simpson, C. Mckillop, and V. Nunn. (2002). Electronic course surveys: Does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education,* 27, 325-337.

Winer, L.R. & Sehgal, R. (2006). Online course evaluation analysis report. Retrieved from http://www.mcgill.ca/files/tls/online_course_evaluation_report.pdf

Young, M., & McCaslin, S. (2013). Student evaluations of faculty revisited: Online versus traditional methods. *Journal of Academic Administration in Higher Education*, 9(2), 11-17.