

Face-to-Face versus Online Course Evaluations: A "Consumer's Guide" to Seven Strategies

Ronald A. Berk

Professor Emeritus of Biostatistics and Measurement
Schools of Education and Nursing
The Johns Hopkins University
Baltimore, MD 21218 USA
rberk1@jhu.edu

Abstract

The research on student rating scales and other measures of teaching effectiveness in face-to-face (F2F) courses has been accumulating for 90 years. With the burgeoning international development of online and blended/hybrid courses over the past decade, the question of what measures to use has challenged directors of distance education programs. Can the traditional F2F scales already in operation be applied to online courses or do all new scales have to be designed? Despite the increasing number of online courses, attention to their evaluation lags far behind that of F2F courses in terms of available measures, quality of measures, and delivery systems. The salient characteristics of F2F and online courses are compared to determine whether they are really different enough to justify separate scales and evaluation systems. Based on a review of the research and current practices, seven concrete measurement options were generated. They are proffered and critiqued as a state-of-the-art "consumer's guide" to the evaluation of online and blended courses and the faculty who teach them.

Keywords: student rating scales, student evaluation of teaching (SET), teaching effectiveness, blended courses, hybrid courses, web-based courses, distance learning, student–instructor interaction, content delivery, evaluation rubrics, technology tools

Introduction

There are nearly 2,000 references on student rating scales used in face-to-face (F2F) courses ([Benton & Cashin, 2012](#)), with the first journal article published 90 years ago ([Freyd, 1923](#)). In higher education there is more research on and experience with student ratings than with all of the other 14 measures of teaching effectiveness combined, including peer, self, administrator, learning outcomes, and teaching portfolio (Berk, 2006, 2013). With all that has been written about student ratings (Arreola, 2007; Berk, 2006; Seldin, 2006), there are three up-to-date reviews ([Benton & Cashin, 2012](#); [Gravestock & Gregor-Greenleaf, 2008](#); [Kite, 2012](#)) that furnish a research perspective from the world of F2F faculty evaluation.

Unfortunately, there has not been nearly the same level of attention given to the rating scales and other measures used for summative decisions about faculty who teach blended/hybrid and online courses and the evaluation of those courses. Given the sizable commitment by colleges and universities to the F2F scales already being used, can they be applied to online courses? Are online courses structured and delivered that differently from F2F courses? Is the use of technology a big factor that should be measured? Do faculty and administrators now need to develop all new measures for the online courses? What are directors of distance education supposed to use?

The purpose of this paper is to clarify the measurement options available to evaluate teaching effectiveness in online courses primarily for faculty employment decisions of contract renewal, merit pay, teaching awards, promotion, and tenure. That information can also be used for course and program evaluation. The first two sections briefly review the status of online courses and the major characteristics of F2F and online courses to determine whether they are really different enough to justify separate measures and evaluation systems. Finally, based on a review of the research and current practices, seven concrete measurement options are described. They are proffered and critiqued as a state-of-the-

art "consumer's guide" to the evaluation of online and blended courses. Selecting the correct options can potentially move formative, summative, and program decisions to a higher level of evaluation practice.

Status of Online Courses

The Pew Research Center's survey of U.S. colleges and universities found that more than 75% offer online courses ([Taylor, Parker, Lenhart, & Moore, 2011](#)). More than 30% of all college enrollments in Fall 2010 were in online courses ([Allen & Seaman, 2011](#)) and nearly 9% of all graduate degrees in 2008 were earned online ([Wei et al., 2009](#)).

The conversion of traditional F2F courses into either blended/hybrid combinations of F2F and online or into fully online courses is increasing at a rapid pace along with enrollments in those courses. Further, there is no sign that these trends are abating nationally ([McCarthy & Samors, 2009](#)) or internationally ([Higher Education Strategy Group, 2011](#)). Distance education in all of its forms is the "course tsunami" of the future. Everyone needs to be prepared.

Unfortunately, evaluation of these online courses and the faculty who teach them lags far behind in terms of available measures, quality of measures, and delivery systems ([Hathorn & Hathorn, 2010](#); [Rothman, Romeo, Brennan, & Mitchell, 2011](#)). Although formative decisions based on student data for course improvement can be conducted by the professor during the course using learning analytics, especially for massive open online courses (MOOCs) ([Bienkowski, Feng, & Means, 2012](#); [Ferguson, 2012](#); [van Barneveld, Arnold, & Campbell, 2012](#)), the overall commitment to online evaluation is lacking. A recent survey of distance learning programs in higher education (Primary Research Group, 2012) in the U.S., Canada, and U.K. found that fewer than 20% of the colleges (15% U.S. and 37.5% Canada and U.K.) have at least one full-time staff person devoted to evaluating the online distance-learning program.

Comparison of Face-to-Face and Online Courses

A brief review of the research on student ratings by [Benton and Cashin \(2012\)](#) and a more extensive review on the evaluation of online courses by [Drouin \(2012\)](#) both came to the same conclusion: *F2F and online courses are more similar than they are different*. They share several key "teaching" factors in common. Lists of some of the common characteristics and the unique characteristics of online courses are given next. Details can be found in the sources cited.

Common Characteristics

[Drouin \(2012\)](#) identified five criteria of "best practices" in online courses that she says could serve as "best practices" in F2F as well for student, peer, and self-ratings. The categories of those criteria are: (1) student-student and student-instructor interactions; (2) instructor support and mentoring; (3) lecture/content delivery quality; (4) course content; and (5) course structure (p. 69). The differences lie in the use of technology related to the delivery of content and social networking tools.

Unique Characteristics of Online Courses

In contrast to Drouin's criteria, [Creasman \(2012\)](#) extracted seven key differences in online courses (p. 2):

- 1) Asynchronous activity, where students can interact with each other and course materials anytime, 24/7;
- 2) Non-linear discussions on message boards and forums, where students can participate in multiple conversations simultaneously;
- 3) Communication primarily via written text;
- 4) Slower communication between instructor and students, primarily via e-mail;
- 5) Greater social contact and time spent by instructor with students on website;
- 6) Greater volume of information and resources available;
- 7) Instructor's roles as a facilitator, "guide on the side," and also co-learner.

So, what do these differences mean in terms of the scales used to measure teaching in online courses? Can these differences be covered on new scales, or should current F2F scales be administered in online courses? This is the problem with which the next section is concerned.

Seven Strategies to Evaluate Teaching Effectiveness in Online Courses

As online courses were being developed and following different models of teaching ([Anderson & Dron, 2011](#); [Creasman, 2012](#); [Peltier, Schibrowsky, & Drago, 2007](#)), existing traditional F2F rating scales were challenged with regard to their application to these courses ([Harrington & Reasons, 2005](#); [Loveland, 2007](#)). The F2F approach seemed efficient since many of those student rating scales were increasingly being administered online at hundreds of institutions. However, these student ratings were just the beginning; *they are a necessary, but not sufficient, source of evidence to evaluate teaching effectiveness in F2F courses* (Berk, 2006, 2013). Other sources must also be used for employment decisions and formative decisions of teaching and course improvement ([Berk, 2005](#)).

This online administration was being executed either by an in-house information technology system or by an outside vendor specializing in online administration, analysis, and score reporting, such as [CollegeNET](#) (*What Do You Think?*), [ConnectEDU](#) (*courseval*) [EvaluationKIT](#) (*Online Course Evaluation and Survey System*), and [IOTA Solutions](#) (*MyClassEvaluation*). The choice of the course management system was crucial in providing the anonymity for students to respond, which could boost response rates ([Oliver & Sautter, 2005](#)). Most of the vendors' programs are compatible with [Blackboard](#), [Moodle](#), [Sakai](#), and other learning management systems.

Despite these online capabilities in place at many colleges and universities, it became apparent that *F2F measures might not address all of the essential components of online teaching* ([Loveland, 2007](#)). This validity issue seriously questioned the actual coverage of instructor behaviors and course characteristics. Perhaps new measures are needed that are tailored to the specific features of those courses.

A review of the research and current practices in evaluating F2F and online courses suggests there are at least seven options for measuring teaching effectiveness. Whatever options are chosen for formative, summative, and program decisions, they must meet the design and technical standards of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education Joint Committee on Standards, 1999), *Personnel Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2009), and *Program Evaluation Standards* (Yarbrough, Shulha, Hopson, & Caruthers, 2011). A critique of these options follows:

- 1) **Instructor-developed scale.** Encourage instructor-developed scales to evaluate online teaching and courses. Some institutions have placed the responsibility for evaluating the online course on the individual instructor or simply neglect the evaluation ([Compura, 2003](#)). Unless instructors are trained in the process of scale construction and score analysis and interpretation for formative or summative decisions, this should not even be considered as a viable option.

Although the technology exists with free online survey providers such as [Zoomerang](#) and several others (see [Wright, 2005](#)) to easily administer online course scales of up to 30 items per scale via e-mail ([Lip, 2008](#)), it is not recommended. Further, after all that has been learned in the evaluation of F2F courses, the complexity of multiple measures, such as student, self, peer, administrator, and mentor rating scales, for formative and summative decisions cannot be handled by each instructor. *Online course assessment should not be the sole responsibility of the instructor.* There are much better ways to do it.

- 2) **Traditional F2F student rating scale.** Use the traditional student rating scale and other measures that are currently in operation for the F2F courses. This is not an uncommon practice for student scales ([Beattie, Spooner, Jordan, Algozzine, & Spooner, 2002](#); [Compura, 2003](#)), but may not be generalizable to self, peer, and other measures.

Studies using the same student rating scale in both types of courses yield comparable ratings on several items, including course and instructor global items, on the *IDEA Student Ratings of Instruction* form ([Benton, Webster, Gross, & Pallett, 2010](#)). Also, there were similar item means, internal consistency reliabilities, and factor structures ([McGhee & Lowell, 2003](#)), and nearly identical overall ratings of the instructor ([Wang & Newlin, 2000](#)).

This continuation of the F2F scale administration to all courses will not capture elements that are unique to each type of course as well as the specific emphases and concentration in delivery methods and technology that may be especially useful for course design and improvement.

Student feedback on the total package, not part of it, would be preferable. The remaining options make that possible.

- 3) **Optional items added to F2F scale.** Use 10 optional items on the F2F traditional scale to measure the online characteristics. This may be the most efficient and cost-effective approach to retain those items in common to both F2F and online courses, but also add items to address the uses of technology and other aspects of teaching that are different for online courses.

Consider a 10-item "optional" subscale tacked onto the "generic" F2F scale. This subscale is usually built into many "home-grown" scales constructed by faculty committees in-house. However, it is also found on most commercially developed scale packages, such as *Online Course Evaluation*, *Student Instructional Report II (SIR II)*, *Course/Instructor Evaluation Questionnaire (CIEQ)*, *IDEA Student Ratings of Instruction*, *Student Evaluation of Educational Quality (SEEQ)*, *Instructional Assessment System (IAS Online)*, and *Purdue Instructor Course Evaluation Service (PICES)*.

In F2F courses, these 10 course-specific items would permit instructors to customize at least a portion of the total scale to the unique characteristics of their respective courses. However, in online and blended/hybrid courses, the 10 items would be a specifically designed standard set of online items measuring technology and other topics, which would be the same for all online courses. Psychometric analysis would have to be conducted on this subscale to establish reliability and validity. In addition, the possible "halo effect" from the common core items ([Madden, Dillon, & Leak, 2010](#)) and other types of response bias (Berk, 2010) should be investigated to estimate the magnitude of overall bias in the 10-item subscale scores.

A major advantage of this option is the *norms available for various score comparisons*. This scale configuration would allow comparisons between F2F and online courses on the same "generic" scale and also between online and blended courses on the 10-item "online" subscale. That feedback would provide valuable information for formative, summative, and program decisions.

- 4) **Revision of F2F scale.** Adapt or revise the current scales to fit the online courses. This is an extension of the preceding strategy when the 10-item add-on subscale is not adequate. If more items are required to cover online teaching, then this may be your option, which was suggested by [Tallent-Runnels et al. \(2006\)](#). The add-on subscales measuring course characteristics and teaching behaviors in an online environment may also be used in blended/hybrid courses. As with option 3, the new items and subscales would require item analysis, reliability, and validity studies, especially bias, to compare their psychometric quality with the rest of the scale.
- 5) **New rating scale.** Develop new rating scales for online courses. This "throw-the-baby-out..." approach may be the *most time-consuming and costly option*, plus it may not be necessary for student rating scales. Building new measures from scratch intended solely for online courses is not without precedent. Several examples at half a dozen institutions for constructing self- and peer-rating scales to fit three models of online teaching (cognitive behaviorist, social constructivist, connectivist) are described by [Drouin \(2012\)](#) using different course rubrics.

Among the 15 sources of evidence that can be used to evaluate F2F teaching effectiveness (Berk, 2013), several may have to be reconfigured for online courses (e.g., external expert ratings, mentor's advice, and student interviews) or created from scratch (e.g., teaching portfolio, administrator ratings, exit/alumni ratings, and employer ratings). While course rubrics can furnish a starting point for designing these measures, considerable attention should be devoted to their psychometric quality.

- 6) **Commercially developed student rating scale.** Use one of the commercially developed online scales already available. Among the seven commercial packages listed in option 3, *only two vendors market scales that are designed expressly for online courses*: (1) *e-SIR*, which is a 41-item scale with half of the items from the *SIR II* by [Educational Testing Service](#); and (2) a 31-item distance learning form from the *IAS Online* at the [Office of Educational Assessment, University of Washington](#). Neither vendor reports any reliability coefficients or validity evidence on its website. These scales should be examined carefully by directors of distance education programs for their content coverage compared to that of the published scales described next. There are major differences to consider.

- 7) **Published student rating scale.** Use available published scales designed for online courses. However, as noted above, there are resources available at other institutions that can assist in adopting, adapting, or developing online scales ([Hathorn & Hathorn, 2010](#); [Hosie, Schibeci, & Backhaus, 2005](#); [MarylandOnline, 2013](#)).

There are also three published student rating scales identified by [Drouin \(2012\)](#) that were designed expressly for online courses, which are worthy of review:

- *Student Evaluation of Online Teaching Effectiveness (SEOTE)* ([Bangert, 2006, 2008](#)), a 26-item Likert-type "agree–disagree" 6-point scale, which is specifically aligned with Chickering and Gamson's (1987) seven principles of effective teaching; a large portion of the items can be used in F2F courses as well; high reliability with four interpretable factors (student–faculty interaction, active learning, time on task, cooperation among students)
- *Student Evaluation of Web-Based Instruction (SEWBI)* ([Stewart, Hong, & Strudler, 2004](#)), a lengthy, very comprehensive 59-item Likert-type "agree–disagree" 5-point scale, which has a very heavy emphasis on the technological aspects of the course and use of multimedia; moderate to high reliability (.75 to .85 for all subscales except technical issues with .92); factor analysis and results reported in seven dimensions (appearance of web pages, hyperlinks and navigation, technical issues, online applications, class procedures and expectations, content delivery, and instructor and peer interaction)
- *Students' Perceptions of Online Courses (SPOC)* ([Rothman et al., 2011](#)), a 25-item Likert-type "agree–disagree" 5-point scale, which covers both general and specific behaviors mostly appropriate for online courses; high reliability with items distributed across six factors (appropriateness of readings/assignments, technological tools, instructor feedback and communication, course organization, clarity of outcomes and requirements, content format); supplements *SIR II* scale administered to F2F courses

Recommendations

Online courses are sprouting up on campuses worldwide. Unfortunately, some institutions have not yet even planted evaluations of those courses. When they are planted, directors of distance education are not sure exactly what the evaluations should look like. The preceding review suggests a state-of-the-art "consumer's guide" to the evaluation of online and blended courses. A few questions need to be answered before picking the most appropriate option.

Should Uniqueness of Online Courses Be Measured?

Despite the similarity of ratings in F2F and online courses with traditional F2F scales and the availability of those scales, the *unique characteristics of online courses should be measured* to furnish a more complete, as opposed to biased, evaluation of those courses and the faculty who teach them. That would eliminate option 2. The issue is how to do that efficiently and cost effectively to produce psychometrically defensible scales.

Which Options Match the Decisions to Be Made?

The decision should drive the choices of evidence. Think carefully about the decision in terms of the time frame, conditions, information needed, and the personnel or program about which the decision will be made. Then select the most appropriate sources for the specific decision.

Start with student rating scales and the options described, then add peer, self, and other sources from among 15 possible sources. Option 5 mentioned a few of those sources and the role of course rubrics in their design. The psychometric quality and interpretation of the sources for formative and summative decisions about online faculty should be guided by the *Personnel Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2009), and online course and program decisions should follow the *Program Evaluation Standards* (Yarbrough et al., 2011).

Should Criterion- or Norm-Referenced Interpretation of Ratings Be Used?

Criterion-referenced ratings are collected and compared *within* each course, but all norm groups and comparisons are *outside* of the course, thereby furnishing a different frame of reference and new information. All of the seven options permit criterion-referenced interpretations of item anchor, item,

subscale, and scale ratings (Berk, 2013); normative comparisons are possible for all except option 1. As noted above, the decision will determine the sources and types of rating interpretations for individual online faculty, administrators at different levels, and curriculum committees for online programs.

Conclusion

The seven options presented in the last section are worthy of consideration. Options 3, 4, 6, and 7 seem to have the greatest potential for student rating scales. Option 5 will probably be necessary for self, peer, and other rating scales to provide a comprehensive assessment of online teaching effectiveness. Those multiple measures are essential to furnish the evidence needed: (1) by faculty for formative decisions to improve their teaching along with learning analytics; (2) by administrators for summative decisions on contract renewal, merit pay, promotion and tenure, and teaching awards; and (3) by directors of distance education for program decisions.

Further research is required to empirically evaluate the efficacy, efficiency, cost-effectiveness, and psychometric soundness of the preceding options as they are implemented. There is no simple solution or generic model that will fit every institution's online course evaluation needs. Similar to F2F applications, appropriate rating scales and other measures of teaching effectiveness must be tailored to the specific online and/or blended courses, faculty, students, culture, and resources at each institution. The seven options that have been suggested in this article provide a few strategies to consider.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education Joint Committee on Standards. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Allen, I. E., & Seaman, J. (2011). *Going the distance: Online education in the United States, 2011*. Babson Park, MA: Babson Survey Research Group. Retrieved from <http://www.onlinelearningsurvey.com/reports/goingthedistance.pdf>
- Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *The International Review of Research in Open and Distance Learning*, 12(3), 80-97. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/890/1663>
- Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system* (3rd ed.). Bolton, MA: Anker.
- Bangert, A. W. (2006). Identifying factors underlying the quality of online teaching effectiveness: An exploratory study. *Journal of Computing in Higher Education*, 17(2), 79-99. doi:10.1007/BF03032699
- Bangert, A. W. (2008). The development and validation of the Student Evaluation of Online Teaching Effectiveness. *Computers in the Schools*, 25(1-2), 25-47. doi:10.1080/07380560802157717
- Beattie, J., Spooner, F., Jordan, L., Algozzine, B., & Spooner, M. (2002). Evaluating instruction in distance learning classes. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 25(2), 124-132. doi:10.1177/088840640202500204
- Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature*. Manhattan, KS: The IDEA Center. Retrieved from http://www.theideacenter.org/sites/default/files/idea-paper_50.pdf
- Benton, S. L., Webster, R., Gross, A. B., & Pallett, W. H. (2010). *An analysis of IDEA student ratings of instruction in traditional versus online courses*. Manhattan, KS: The IDEA Center. Retrieved from <http://www.theideacenter.org/sites/default/files/Technical%20Report15pdf.pdf>
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62. Retrieved from <http://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf>

- Berk, R. A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, VA: Stylus.
- Berk, R. A. (2010). The secret to the "best" ratings from any evaluation scale. *The Journal of Faculty Development*, 24(1), 37-39.
- Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions*. Sterling, VA: Stylus.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: Office of Educational Technology, U.S. Department of Education. Retrieved from http://www.evidenceframework.org/wp-content/uploads/2012/04/EDM-LA-Brief-Draft_4_10_12c.pdf
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 39(7), 3-7.
- Compura, D. (2003). Current trends in distance education: An administrative model. *Online Journal of Distance Learning Administration*, 6(2). Retrieved from <http://www.westga.edu/~distance/ojdla/summer62/compura62.html>
- Creasman, P. A. (2012). *Considerations in online course design*. Manhattan, KS: The IDEA Center. Retrieved from http://www.theideacenter.org/sites/default/files/idea_paper_52.pdf
- Drouin, M. (2012). What's the story on evaluations of online teaching? In M. E. Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators* (pp. 60-70). Washington, DC: Society for the Teaching of Psychology. Retrieved from <http://www.teachpsych.org/Resources/Documents/ebooks/evals2012.pdf>
- Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges*. Milton Keynes, UK: Knowledge Media Institute, The Open University. Retrieved from <http://www.kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf>
- Freyd, M. (1923). A graphic rating scale for teachers. *The Journal of Educational Research*, 8(5), 433-439. Available from JSTOR database. (27524960)
- Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Toronto, Canada: Higher Education Quality Council of Ontario. Retrieved from http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations_Research,%20Models%20and%20Trends.pdf
- 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>
- Hathorn, L., & Hathorn, J. (2010). Evaluation of online course websites: Is teaching online a tug-of-war? *Journal of Educational Computing Research*, 42(2), 197-217. doi:10.2190/EC.42.2.d
- Higher Education Strategy Group. (2011). *National strategy for higher education to 2030: Report of the Strategy Group*. Dublin, Ireland: Department of Education and Skills. Retrieved from <http://www.education.ie/en/Publications/Policy-Reports/National-Strategy-for-Higher-Education-2030.pdf>
- Hosie, R., Schibeci, R., & Backhaus, A. (2005). A framework and checklists for evaluating online learning in higher education. *Assessment & Evaluation in Higher Education*, 35(5), 539-553. doi:10.1080/02602930500187097
- Joint Committee on Standards for Educational Evaluation. (2009). *The personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed.). Thousand Oaks, CA: Corwin.
- Kite, M. E. (Ed.). (2012). *Effective evaluation of teaching: A guide for faculty and administrators*. Washington, DC: Society for the Teaching of Psychology. <http://www.teachpsych.org/Resources/Documents/ebooks/evals2012.pdf>

- Lip, P. C. H. (2008). Evaluating teaching and learning from students' perspectives in their classroom through easy-to-use online surveys. *International Journal of Cyber Society and Education*, 1(1), 33-48. Retrieved from <http://www.academic-journals.org/ojs2/index.php/IJCSE/article/download/451/4>
- Loveland, K. A. (2007). Student evaluation of teaching (SET) in web-based classes: Preliminary findings and a call for further research. *The Journal of Educators Online*, 4(2), 1-18. Retrieved from <http://www.thejeo.com/Volume4Number2/Loveland%20Final.pdf>
- Madden, T. J., Dillon, W. R., & Leak, R. L. (2010). Students' evaluation of teaching: Concerns of item diagnosticity. *Journal of Marketing Education*, 32(3), 264-274. doi:10.1177/0273475310377759
- MarylandOnline. (2013). *Higher ed program > Rubric | Quality Matters program*. Retrieved from <http://www.qualitymatters.org/rubric/>
- McCarthy, S. A., & Samors, R. J. (2009). *Online learning as a strategic asset. Volume I: A resource for campus leaders*. Washington, DC: Association of Public and Land-grant Universities. Retrieved from <http://www.aplu.org/NetCommunity/Document.Doc?id=1877>
- McGhee, D. E., & Lowell, N. (2003). Psychometric properties of student ratings of instruction in online and on-campus courses. *New Directions for Teaching and Learning*, 96, 39-48. doi:10.1002/tl.121
- Oliver, R. L., & Sautter, E. P. (2005). Using course management systems to enhance the value of student evaluations of teaching. *Journal of Education for Business*, 80(4), 231-234. doi:10.3200/JOEB.80.4.231-234
- Peltier, J. W., Schibrowsky, J. A., & Drago, W. (2007). The interdependence of the factors influencing the perceived quality of the online learning experience: A causal model. *Journal of Marketing Education*, 29(2), 140-153. doi:10.1177/0273475307302016
- Primary Research Group. (2012). *The survey of distance learning programs in higher education, 2012-13 edition*. New York, NY: Author.
- Rothman, T., Romeo, L., Brennan, M., & Mitchell, D. (2011). Criteria for assessing student satisfaction with online courses. *International Journal for e-Learning Security*, 1(1-2), 27-32. Retrieved from <http://www.infonomics-society.org/IJeLS/Criteria%20for%20Assessing%20Student%20Satisfaction%20with%20Online%20Courses.pdf>
- Seldin, P. (Ed.). (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*. Bolton, MA: Anker.
- Stewart, I., Hong, E., & Strudler, N. (2004). Development and validation of an instrument for student evaluation of the quality of web-based instruction. *American Journal of Distance Education*, 18(3), 131-150. doi:10.1207/s15389286ajde1803_2
- Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research*, 76(1), 93-135. doi:10.3102/00346543076001093
- Taylor, P., Parker, K., Lenhart, A., & Moore, K. (2011). *The digital revolution and higher education*. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://www.pewinternet.org/~media/Files/Reports/2011/PIP-Online-Learning.pdf>
- van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). *Analytics in higher education: Establishing a common language*. Boulder, CO: EDUCAUSE. Retrieved from <http://net.educause.edu/ir/library/pdf/ELI3026.pdf>
- Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, 92(1), 137-143. doi:10.1037/0022-0663.92.1.137
- Wei, C. C., Berkner, L., He, S., Lew, S., Cominole, M., Siegel, P., & Griffith, J. (2009). *2007-08 National Postsecondary Student Aid Study (NPSAS: 08): Student financial aid estimates for 2007-08, First look*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://www.nces.ed.gov/pubs2009/2009166.pdf>

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3). doi:10.1111/j.1083-6101.2005.tb00259.x

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.



This work is published under a Creative Commons Attribution-Non-Commercial-Share-Alike License

For details please go to: <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>