

Validity of Response Rates: Student Comments vs. Rating Scale Format

An Annotated Bibliography

Compiled by Lisa LeVasseur, Institutional Effectiveness, WMU

Barkhi, R., & Williams, P. (2010). The impact of electronic media on faculty evaluation. *Assessment and Evaluation in Higher Education*, 35(2), 241.

Abstract summary: With the proliferation of computer networks and the increased use of Internet-based applications, many forms of social interactions now take place in an on-line context through Computer-Mediated Communication (CMC). Many universities are now reaping the benefits of using CMC applications to collect data on *student evaluations* of faculty, rather than using paper-based surveys in Face-To-Face (FTF) classroom settings. While the relative merits of CMC versus FTF *student evaluations* have been researched extensively, there is limited research published about the ways students respond to the questions from either mode of data collection. This paper reports on a research study to analyse the communication differences between student scores from FTF *student evaluations* and CMC evaluation questions from end of semester evaluations from a university in the Middle East region. In addition to the questions about communication mode differences between two evaluation questions, several demographic variables were measured to determine any interaction effects. The results of our study suggest that the type of communication channel mitigates the responses that students make on CMC evaluations vis-a-vis FTF evaluations of faculty. In particular, even though there were significant differences found at the aggregate level between CMC and FTF evaluations, when the course and instructor are controlled for, there were no significant differences reported. In addition, several differences were noted depending on the type and level of the course being studied. Also, we found that students are more likely to express more extreme responses to scale questions in CMC than FTF evaluations. Administrators should consider these potential differences when implementing on-line evaluation systems.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *The American Psychologist*, 52(11), 1198-1208.

Abstract summary: Many colleges and universities have adopted the use of student ratings of instruction as one (often the most influential) measure of instructional effectiveness. In this article, the authors present evidence that although effective instruction may be multidimensional, student ratings of instruction measure general instructional skill, which is a composite of three subskills: delivering instruction, facilitating interactions, and evaluating student learning. The authors subsequently report the results of a metaanalysis of the multisection validity studies that indicate that student ratings are moderately valid; however, administrative, instructor, and course characteristics influence student ratings of instruction.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *The American Psychologist*, 52(11), 1182-1186.

Abstract summary: The validity of student rating measures of instructional quality was severely questioned in the 1970s. By the early 1980s, however, most expert opinion viewed student rating measures as valid and as worthy of widespread use. In retrospect, older discriminant-validity concerns were not so much resolved as they were displaced from research attention by accumulating evidence for convergent validity. This article introduces a Current Issues section that gives new attention to validity concerns associated with student ratings. The section's 4 articles deal, respectively, with (a) conceptual structure (are student ratings unidimensional or multidimensional?), (b) convergent validity (how well do ratings correlate with other indicators of effective teaching?), (c) discriminant validity (are ratings influenced by factors other than teaching effectiveness?), and (d) consequential validity (are ratings used effectively in personnel development and evaluation?). Although all 4 articles favor the use of ratings, they disagree on controversial points associated with interpretation and use of ratings data.

Hendry, G. D., & Baur, L. A. (1998). Monitoring evaluation of a paediatrics and child health course: Factors in implementing course improvement. *Assessment and Evaluation in Higher Education*, 23(2), 179-189.

Abstract summary: Although questionnaire rating scales are a commonly used method of course evaluation in higher education they have several limitations. When a course remains unchanged, despite student feedback, both students and teachers may see evaluation as a pointless exercise. In comparison to questionnaires, focus group methods are increasingly being used in health and educational programme evaluation. This paper reports on an evaluation process in a new undergraduate Paediatrics and Child Health course and evaluation methods that were found to be most effective are discussed. The successful implementation of course improvements depended on several key factors including an established aim, collecting feedback through student focus groups, responding to students about changes planned, and strong leadership for the process of change.

Jones, J. E. (1990). Don't smile about smile sheets. *Training and Development Journal*, 44(12), 19.

Abstract summary: Most training courses include end-of-course questionnaires that ask participants to rate various aspects of the experience. In the training industry, these instruments are referred to as smile sheets. The most common reasons for using them are: 1. Training sponsors want them. 2. The training staff wants to know how it performed. 3. The ratings look valid. Unfortunately, the usefulness of these questionnaires is severely limited. Twenty-six specific limitations are detailed, along with 7 suggestions on how to handle the limitations. It is concluded that anyone who wishes to gather complete information on the results of training should rely on more than one method and more than one data source.

Marsh, H. W., & Roche L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *The American Psychologist*, 52(11), 1187-1197.

Abstract summary: This article reviews research indicating that, under appropriate conditions, students' evaluations of teaching (SETs) are (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases (e.g., grading leniency, class size, workload, prior subject interest); and (f) useful in improving teaching effectiveness when SETs are coupled with appropriate consultation. The authors recommend rejecting a narrow criterion-related approach to validity and adopting a broad construct-validation approach, recognizing that effective teaching and SETs that reflect teaching effectiveness are multidimensional; no single criterion of effective teaching is sufficient; and tentative interpretations of relations with validity criteria and potential biases should be evaluated critically in different contexts, in relation to multiple criteria of effective teaching, theory, and existing knowledge.

Timmerman, T. (2008). On the validity of ratemyprofessors.com. *Journal of Education for Business*, 84(1), 55-61.

Abstract summary: The purpose of this study was to address some of the most common questions and concerns regarding RateMyProfessors.com (RMP). Data from 5 different universities and 1,167 faculty members showed that (a) the ratings are not dominated by griping, (b) the summary evaluation correlates highly with summary evaluations from an official university evaluation, (c) substantive relations are generally the same when only a single rating has been provided, (d) the relation between RMP Easiness and RMP Quality is partially explained by the fact that learning is associated with perceived easiness, and (e) the substantive findings generalize to business faculty in different universities. The author discusses the possible value of RMP without endorsing its unlimited use for administrative purposes.

Trout, P. A. (1997). What the numbers mean. *Change*, 29(5), 24-30.

Abstract summary: The context for numerical student evaluations of courses, which are used in about 80% of postsecondary institutions, are discussed. Trout examines the reliability and validity of the scores.

Vicente, P., & Reis, E. (2010). Using questionnaire design to fight nonresponse bias in web surveys. *Social Science Computer Review*, 28(2), 251.

Abstract summary: The technical potential of the Internet offers survey researchers a wide range of possibilities for web surveys in terms of questionnaire design; however, the abuse of technical facilities can detract respondents from cooperating rather than motivating them. Within the web survey methodology literature, many contributions can

be found on how to write a "good" questionnaire. The outcomes are however scattered and researchers and practitioners may find it difficult to obtain an overall picture. The article reviews the latest empirical research on how questionnaire characteristics affect *response rates*. The article is divided into three main sections: an introduction where the various forms of nonresponse in web surveys are described; a second section presenting questionnaire features affecting nonresponse--general structure, length, disclosure of survey progress, visual presentation, interactivity, and question/response format--and a final section that summarizes the options in terms of questionnaire design and its implications for nonresponse rate.