Internet and Higher Education 14 (2011) 236–241

Contents lists available at ScienceDirect

Internet and Higher Education

A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments

Charles Dziuban *, Patsy Moskal [1]

Research Initiative for Teaching Effectiveness, P.O. Box 162810, Orlando, FL 32816-2810, University of Central Florida, USA

ARTICLE INFO

Article history:
Accepted 18 May 2011
Available online 26 May 2011

Keywords:
Student rating of instruction
Online learning
Blended learning
Factor analysis
Student agency

ABSTRACT

The authors compared the underlying student response patterns to an end-of-course rating instrument for large student samples in online, blended and face-to-face courses. For each modality, the solution produced a single factor that accounted for approximately 70% of the variance. The correlations among the factors across the class formats showed that they were identical. The authors concluded that course modality does not impact the dimensionality by which students evaluate their course experiences. The inability to verify multiple dimensions for student evaluation of instruction implies that the boundaries of a typical course are beginning to dissipate. As a result, the authors concluded that end-of-course evaluations now involve a much more complex network of interactions.

1. Introduction

One of the most important impacts of educational technology is that it affords students increased access to contemporary higher education. In part, that access comes from an expanding array of course modalities that students are able to incorporate into their educational programs. A list of emerging formats includes: fully online; blended; face-to-face; video streaming; and courses enhanced with Web 2.0, virtual reality, social networks, mobile devices, and cloud computing. Of these modalities, online and blended courses appear to be the most widespread, although the vast majority of offerings still come under the label "face-to-face". However, most of those "traditional" courses are undergoing some form of enhancement through a number of technological innovations. These opportunities create a rich and varied educational landscape for students to obtain information, experience learning, interact with their peers and instructors, and engage in campus-wide co-curriculum. In many respects, expanded class formats comprise a proactive response to the population's need for educational flexibility and responsiveness.

This unbundled environment, where students have opportunities for comparable learning experiences, whether they are far from campus, near campus, or on campus, transforms colleges and universities into an outreach orientation that alters role expectations for instructors and those they teach (Dziuban, Hartman, Cavanagh, & Moskal, 2010). As a result, students have a much stronger sense of agency in their education, largely because multiple options allow them to assemble personal geographies for how they navigate their learning environment (Harmon, 2004).

Because of the increasingly complex educational landscape, this expanded voice manifests in a greatly altered assessment of the learning experience (Dziuban, Moskal, Brophy-Ellison, & Shea, 2007). This phenomenon influences end-of-course student evaluations, the historical gold standard in higher education, giving them more visibility in the technology-mediated academy. Students publish their own evaluations of instructors making them available through a number of channels, including student government and social organizations. Most campuses have informal social networks where instructors' organic reputations evolve through a continuous student conversation. Recently, the emergence of ratemyprofessors.com takes the expression *student satisfaction* to a broader level (RateMyProfessors, 2011). For better or worse, the website creates a forum for students to evaluate a course and its instructor, and communicate to an audience of considerable size (Coladarci & Kornfield, 2007). Course evaluations radiate to spaces such as Facebook and instant messaging. Today, students are able to tweet followers giving them real-time accounts of what transpires in their classes. The website YouTube contains many video clips of instructors in the act of teaching (YouTube, 2010).

2. Review of literature

2.1. Research on SEI

Examining the literature of higher education, one is struck by the pure volume of research on student evaluation of instruction (SEI). Even a casual review of that literature will underscore the perceived importance of this topic with literally hundreds of studies in discipline

* Corresponding author. Tel.: +1 407 823 5478; fax: +1 407 823 6580.
E-mail addresses: Charles.Dziuban@ucf.edu (C. Dziuban), Patsy.Moskal@ucf.edu (P. Moskal).
[1] Tel.: +1 407 823 5478; fax: +1 407 823 6580.

1096-7516/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.iheduc.2011.05.003

specific, pedagogical, psychological, and measurement journals. Virtually every institution provides a protocol through which students express their satisfaction, dissatisfaction, or ambivalence with their educational experiences. Most often this process takes on the end-of-course rating procedure where summary data serve as the primary feedback mechanism for faculty members and administrators. Ostensibly, the resulting information serves to help instructors improve teaching and provides summative information about instructional effectiveness. In many instances, these student ratings become high stakes issues for faculty members because they contribute to end-of-year evaluation portfolios, promotion status, and salary decisions and they form the basis for teaching awards.

2.2. The foundational work in SEI

Early research in SEI spawned a lively debate between two opposing camps: The first group saw little added value in the process and felt that it constituted a disruptive influence on the learning process (Adams, 1997; Altschuler, 2001; Eiszler, 2002; Greenwald & Gilmore, 1997; Hoyt, 1977). During this period Kolitch and Dean (1999) conducted a particularly interesting critical content analysis for a large number of end-of-course survey instruments concluding that the items particularized a transmission of information model for teaching, and further arguing that such an approach too narrowly defined the parameters of an effective course. The authors suggested alternative items that reflected what they called an engaged critical model of teaching; for instance: As a result of this course have you done anything to improve your community?

Conversely, a group of investigators contended that the process reflected a reliable and valid index of a professor's effectiveness if those evaluations represented multiple perspectives. They argued that students possess the wisdom and experience to evaluate teaching accurately and that the process does not represent a popularity contest caused by lenient grading. Marsh and Roche (1997) concluded that student ratings are reliable, stable, and multidimensional, contending that such data exhibit validity against a variety of teaching effectiveness indicators. Further, they contended that student ratings are unaffected by a number of potentially biasing factors and they can be useful for improving instruction with effective consultation. Felder (1992), another protagonist for student ratings, argued that SEI reflects a reliable and valid index of a professor's effectiveness as well, especially when representing multiple constituencies. He sought to dispel several commonly held myths about student ratings of their learning experiences. In Felder's judgment, certain contentions were particularly invalid: that faculty receive high ratings as a result of lenient grading, that student ratings were popularity contests, and that students lack the experience and wisdom to evaluate teaching. When reviewing this early literature, however, it becomes important to note that virtually all of this research was conducted in the face-to-face environment as innovations such as online and blended learning were yet to appear on the educational scene.

2.3. The debate over dimensionality

Interestingly, that period represented the “hey day” of factor analytic techniques where investigators sought to identify underlying latent dimensions found in student evaluation of instruction. By definition, these factors cannot be observed directly, but must be inferred from some combination of the original variables. In one sense, factor analysis identifies surrogate elements for student rating instruments that clarify understanding of the framework students use to evaluate their courses and instructors. Of course, these procedures have their controversies as well.

The argument framed itself this way: student ratings are complex and require multiple dimensions to adequately describe students' evaluation of the learning environments. The opposing position argued that student ratings are one dimensional, indexing a general teaching factor. In early factor studies, Feldman (1976) offered 20 components by which effective teaching should be assessed and ordered them into three higher order categories: presentation, facilitation, and regulation. Marsh and Roche (1997) proposed a ten factor model for teaching assessment: learning value, instructor enthusiasm, organization and clarity, group interaction, individual rapport, breadth of coverage, examinations and grading, assignments and readings, workload, and difficulty. Kim, Damewood, and Hodge (2000) investigated the affective aspect of student evaluation of teaching, identifying several additional components: demonstrates enthusiasm, encourages student motivation, encourages student discussions, is open to constructive criticism, provides assistance outside of class, encourages students to ask for help, is considerate of students, generates equality among students, respects students, and demonstrates a positive attitude for the course and the students. Using methods of structural equation modeling, Shelvin, Banyard, Davies, and Griffiths (2000) found two major dimensions that define student evaluations: lecture and stability of the relationship, which is heavily mediated by an interactive instructor charisma dimension. Linn, Centra, and Tucker (1975), in an important psychometric study, offered a cautionary note about factoring student ratings of instruction. They contended that many studies ignored the between and within instructor covariance among factors, but found that when they made the comparison the total group solution provided an acceptable fit to the between and within covariance matrices.

Other early investigators disagreed with multidimensional theories for explaining student rating of instruction. Greenwald and Gilmore (1997), Alemami and d'Apollonia (1991), and McKeachie (1997) contended that student ratings represented a single global “G” factor. This one dimensional theory opines that one need not be concerned with multiple dimensions because an overall perception is the driving force in the way students evaluate their courses. Students simply didn't think about nor respond to those multiple frames of reference.

2.4. Online learning

The web-based learning environment forces investigators to reexamine many of their assumptions about higher education: access, learning effectiveness, student satisfaction, faculty satisfaction, and return on investment (scale), for instance (Moore, 2005). Technology-enhanced teaching and learning has given rise to a raft of studies in which statistical hypothesis tests were used as the decision metric for deciding whether or not online course modalities impact student learning outcomes. The most noteworthy of these being the now famous “no significant difference” phenomenon (Russell, 2001) where online and face-to-face modalities constitute quasi-treatment effects in the experimental design sense. Recently, the United States Office of Education funded a comprehensive study of the impact of online instruction on learning effectiveness (U.S. Department of Education, 2009). The meta-analysis found that, on average, students in online learning conditions performed better than those receiving face-to-face instruction. However, the authors go on to conclude that their calculated treatment effects are larger for those modalities that blend elements of online and face-to-face learning and that the results may reflect confounds through additional learning time and instructional elements. The study generated controversy. Jaggers and Bailey (2010) contended that “The Department of Education report does not present evidence that fully online delivery produces superior learning outcomes for typical college courses, particularly among low-income and academically underprepared students” (p. 1). Figlio, Rush, and Yin (2010) reported the results of a comparative study in which they were able to randomly assign students to face-to-face or what they term “Internet instruction” in a microeconomics class. However, their study more accurately indexed the impact of streamed lecture capture

rather than asynchronous online learning management. These authors contended "Counter to the conclusions drawn by a recent U. S. Department of Education meta-analysis of non-experimental analyses of Internet instruction in higher education, we find modest evidence that live-only instruction dominates Internet instruction" (p. 1). These studies on learning effectiveness attributable to technology-mediated instruction underscore the intense interest in understanding these emerging learning environments and reframing instruction to accommodate the potential and challenges that multiple course modes offer.

2.5. Factor analytic studies of the SEI in online learning

Recently, a few researchers turned their attention to student evaluation factor identification in the online environment. Stewart, Hong, and Strudler (2004) developed and validated an assessment instrument specifically targeted to web-based learning by analyzing approximately 1400 responses for multiple institutions and instructors using principal component and maximum likelihood methods. After encountering difficulties with the maximum likelihood approach, they used their principal components solution to identify seven factors that underlie student evaluation of online learning: web page appearance; hyperlinks and navigation; technical issues; online application; class procedures and expectations; content delivery; and instructor and peer interaction.

Bangert (2006) studied student evaluation of instruction (n = 498), co-mingling online and blended modalities with undergraduate and graduate students. His principal component solution identified four dimensions for these learning environments: student/faculty interaction, active learning, time on task, and cooperation among students. In a later study with a larger sample (n = 807), he used both explanatory and confirmatory factor procedures verifying the same four dimensions found in his previous study (Bangert, 2008).

Using classification and regression trees (CART), Wang, Dziuban, Cook, and Moskal (2009) developed a set of decision rules for determining what characteristics impact the odds ratio for students assigning an overall rating of excellent to a course. With a large sample (n = 588,575) of student responses to the course evaluation instrument (Appendix A), they used the overall course evaluation item as the dependent variable. The predictors were the remaining 15 items on the student questionnaire, college membership, and course level (lower undergraduate, upper undergraduate, and graduate). They found three predictive rules that excluded course level and college membership. Only other items on the rating scale contributed to overall instructor rating: the instructors' facilitation of learning, communication of ideas and information, organization of the course, and assessment of student progress. Later, they showed that the rules held for online, blended, and face-to-face courses as well.

3. Materials and methods

3.1. The present study

Over the past decade, fully online and blended (reduced seat time combined with online instruction) courses and programs have become a systemic initiative for this university. By providing alternatives to the face-to-face modality, these technology-enhanced courses allow the university to expand access while maintaining quality in an environment of rapid expansion and deteriorating economic conditions. In fact, in the 2010–2011 academic year, approximately 30% of the university's student credit hour production came from some form of technology-enhanced course modality, accommodating virtually all of the institution's growth.

For many years, the university has had in place the Student Evaluation of Instruction rating scale instrument that consists of 16 five-point, Likert scale items (Appendix A). Currently, students complete that instrument online at the conclusion of courses in all modalities. Over the past years, this instrument has produced a very large data set of student end-of-course evaluations for face-to-face, online, and blended courses. The purpose of this study was to compare the underlying factor patterns of student responses to their courses in those three modalities to determine if they are multidimensional or single-dimensional in each modality and if they are the same or different in each modality.

The question becomes: *Is the identical student rating instrument measuring the same or different underlying teaching and learning constructs, depending on the modality in which the course is offered*? One obvious question is why pursue such a study? The answer comes from the psychometric literature on factor invariance (Harris & Harris, 1971). This body of research shows that the same items administered in different contexts can be measuring quite different constructs. Therefore, if this is true, one might expect the factors to vary depending on the course context in which they are administered. However, if the class context is not impacting the dimensionality of the instrument, one might infer some similarity in the way that students respond to their educational experiences across class modes.

To answer these questions the authors collected a large sample (n = 1,124,979) of student end-of-course evaluation responses for the instrument contained in Appendix A. The sample represented undergraduate and graduate courses in online (n = 80,316), blended (n = 48,327), and face-to-face (n = 996,426) formats. To identify the underlying dimensionality of the instrument in the three modalities, the authors computed analogs to common factors. Prior to the analysis procedure, the domain sampling properties of each covariance matrix were assessed using Kaiser's measure of sampling adequacy-MSA (Kaiser & Rice, 1974). The index capitalizes on a Guttman theorem pointing out that if one has an adequate sample of items from a specified domain, then the inverse of the correlation matrix under consideration should approach a diagonal (Guttman, 1954). As MSA approaches 1.0, it gives increasing indication of a better psychometric sample, i.e., the item sample from the specified domain. Dziuban and Shirkey (1974) recommended that investigators use this procedure before they attempt any factoring methods with their data to guard against analysis of noise.

3.2. Image analysis

To assess the comparative dimensionality of student end-of-course ratings for online, blended, and face to courses, the authors used Guttman's image analysis (Guttman, 1953). Mulaik describes the procedure by considering one's dataset in two parts: that which can be predicted from the rest of the variables in the set (the image) and that which is not predictable from those variables (the anti-image) (Mulaik, 1972). This was Guttman's approach for dealing with the basic indeterminacy of the factor analytic model. The image matrix, on which the procedure basis itself, is formed by computing the predicted standardized score (unit normal) for each individual regressed from the remaining variables and computing the covariances among all variables. One should note that most of the current authors of online "factor analysis" studies use principal components as the basis of their analysis. Technically, that procedure is not factor analysis in the Thurstone–Spearman sense and includes the unique and the common part of the data. For this study, image components were retained according to the eigenvalues of the correlation matrix greater than one. The intent was to rotate the original solution according to the Promax procedure (Hendrickson & White, 1964).

4. Results

Table 1 contains the summary characteristics of the datasets for the online, blended, and face-to-face courses. The measures of

Table 1
Data set summary characteristics.

	Online	Blended	Face-to-face
n	80,316	48,327	996,426
Eigenvalue>1	11.12	11.37	11.19
MSA	.97	.96	.95

sampling adequacy for each dataset approached 1.0, which according to Kaiser's calibration of the index, indicates that the 16 items of the instrument comprised an excellent domain sample for each of the modalities, and, therefore, the derived factors would be a good portrayal of the item population. This is the psychometric sampling issue that is a noteworthy finding because if the items were interacting with the modalities, one might reasonably expect some variation of the MSAs. Secondly, for each modality there was only one eigenvalue greater than one, suggesting a single component solution that in each case accounted for approximately 70% of the total system variance in blended, online, and face-to-face courses. Because the total of the eigenvalues in each of the standardized matrices was 16 and the first value in each matrix was over 11 with the second ordered value well below 1.0 in each mode, the resulting one factor solution would mitigate any possibility of component rotation.

Table 2 presents the eigenvalues for each correlation matrix for the modalities showing the dramatic drop off after the first entry.

Table 3 contains the derived image components for each of the three modalities. For each solution, the one retained component produced pattern coefficients that would make all items salient on the single dimension. Interestingly, when Thurstone (1935) devised the theory of common factor analysis he proposed the notion of simple structure. That is, the ideal final pattern matrix would be comprised of "loadings" that would be either ones or zeros, so that there would be no doubt as to which variables contributed to the factors. Unfortunately, factor methods, depending on the data at hand, only permit approximations to simple structure, so the investigators use rules of thumb to decide which variables to assign to which factors. In most published studies, pattern coefficients absolutely greater than .3 or .4 have been the decision rule basis. Interestingly, all of the pattern coefficients in Table 2 are double the customary .3 value and most of them are double the .4 level. Another way to interpret these values is to consider them approximations to correlations of variables with the factors large by most standards. This single evaluation dimension finding is further verified when one considers the alpha reliability coefficients in the three modalities: online = .91, blended = .92, and face-to-face = .93. Table 4 further confirms the single factor similarity. All correlations among the factors approach 1.0, accounting for virtually 100% percent of the variance among them. Therefore, the finding of this study is that when students respond to the end-of-course evaluation instruments for online, blended, and face-to-face courses, the dimensionality in each mode is one, the underlying factors are identical, and they do not differentiate the instructional idiosyncrasies found in the three modalities.

Table 2
Eigenvalues for the item correlation matrices in the three modalities*.

Online	Blended	Face-to-face
11.55	11.37	11.20
.64	.60	.64
.58	.49	.55
.44	.49	.49
.40	.38	.41
.35	.36	.37
.31	.34	.34
.27	.28	.30
.25	.26	.27
.23	.24	.25
.20	.23	.23
.19	.22	.22
.17	.20	.22
.16	.19	.21
.16	.19	.20
.13	.17	.18

*Rounded.

Table 3
Image pattern coefficients for student evaluation of instruction for three course modes*.

		Online	Blended	Face-to-face
1	Feedback concerning performance	82	77	77
2	Instructor interest	86	83	82
3	Use of class time	76	80	80
4	Instructor's overall organization	83	83	83
5	Continuity	77	82	81
6	Pace of the course	78	80	80
7	Instructor's assessment of progress	85	84	82
8	Texts and learning materials	73	72	68
9	Description of objectives and assignments	83	82	83
10	Communication of ideas and information	88	87	86
11	Expectations for performance	87	86	85
12	Availability to assist	80	78	74
13	Respect and concern for students	84	80	78
14	Stimulation of interest	87	84	83
15	Facilitation of learning	87	86	85
16	Overall assessment	92	89	89

*Decimals omitted.

5. Discussion

The findings of this study flow counter to much of the research literature in online and blended learning that emphasizes differences in the modes produce when compared to face-to-face learning environments (Ferguson & Tryjankowski, 2009; U.S. Department of Education, 2009). Further, the results contradict many investigators, who argue that there are multiple dimensions by which students evaluate the quality of their educational experience. This study differs from previous research in a number of other ways. First, the sample size of over one million student responses moves the study from one of estimation to one of data mining. In fact, is using the small letter *n* appropriate when specifying the number of subjects in this case? What does a standard error mean with a million observations? In addition, attempts were made to avoid the shortcomings of the principal component method which does not require a full rank matrix and does not partial out the unique parts of the data. In addition, the investigators attempted to eliminate the often encountered computation problems in the maximum likelihood procedure and the issues encountered in using statistical tests for the number of factors, which is particularly troublesome with large sample sizes (Kaiser, 1976).

These results suggest that students do not consider course mode an important element when defining the dimensions by which they evaluate their educational experience. Intuitively, this seems wrong when one considers items on the rating scale that appear to be intended for the face-to-face environment. The single dimensionality of the patterns found here point to the fact that students pay much more attention to the overall educational experience and less attention to the individual aspects of a course identified in the rating questions. One explanation for this might be that they contextualize the individual

Table 4
Correlations among the components among online, blended, and face-to-face modalities*.

	Blended	Face-to-face
Online	95	99
Blended		98

*Decimals omitted.

items for the online and blended environments and do not use their face-to-face experiences as a standard for comparison. Secondly, it may be that students in technology-mediated courses build their own personal environment for learning and the structure of the course has less to do with how they evaluate their experiences. This seems interesting since much of the onus for learning falls on the student in the constructivist nature of online and blended courses (Mayes, 2001). Possibly, the evaluation in those environments may be in part an assessment of their personal learning strategies in addition to course and instructor.

6. Conclusion

The reader should be clear that by demonstrating a one-component structure for students' evaluation of online, blended and face-to-face courses, the authors are not suggesting that the satisfaction levels at which they rate those modes are equivalent. They contend that students use the same generalized criterion in each case. However, what has been shown is that mode is not an effective predictor of success or withdrawal in courses. The strongest predictor of success is in previous academic performance (Dziuban, 2011). Historically, students who have done well in courses do well in any mode; a course is a course. Combined with the results presented in this study, a reasonable conclusion seems to be that course modality is not a treatment in the experimental or even quasi-experimental case. Students react generally to the course, the content, the instructor, the learning climate, and themselves. Therefore, the authors are unable to confirm multidimensional theories of student evaluation of instruction.

One final comment about end-of-course evaluation surveys in general; with increased access, certain opportunity costs arise. Shirky (2008) makes a compelling argument that the boundaries of what has been called the "class" are disappearing. For students, even those considered traditional, the concept of a self-contained course becomes much less clearly defined as they interact with the instructor, with each other, and the outside world. The class for many contemporary students is an increasingly complex network of interactions. Recently in discussing generational difference a student said to the authors, "You will learn by reading a book or the manual. We will learn through interaction with each other and the Internet." (Hartman, Dziuban, & Brophy-Ellison, 2007). If this is true, then end-of-course evaluations offer very little opportunity for interaction. Perhaps one should consider new instructional evaluation protocols that are not summative and after the fact, but instead those that are more reflective and interactive. The opportunity cost involved would require a much greater investment of time and resources, but good news is that such an approach would increase the student voice in their learning and support a more collaborative educational environment.

Appendix A

Student perception of instruction items*.

1	Feedback concerning your performance in this course was
2	The instructor's interest in your learning was
3	Use of class time was
4	The instruction's overall organization of the course was
5	Continuity from one class meeting to the next was
6	The pace of the course was
7	The instructor's assessment of your progress in the course was
8	The texts and supplemental learning materials used in the course were
9	Description of course objectives and assignments
10	Communication of ideas and information
11	Expression of expectations for performance
12	Availability to assist students in or outside of class
13	Respect and concern for students
14	Stimulation of interest in the course
15	Facilitation of learning
16	Overall assessment of instructor

*Rated on a 5-point Likert Scale: excellent, very good, good, fair, poor.

References

Adams, J. V. (1997). Student evaluations: The ratings game. *Inquiry*, *1*(2), 10–16.

Alemami, P. C., & d'Apollonia, S. (1991). Multidimensional students' evaluation of teaching effectiveness — generalizability of "N=1" research: Comment on Marsh (1991). *Journal of Educational Psychology*, *83*(3), 411–415.

Altschuler, G. C. (2001). Let me edutain you. Retrieved November 30, 2010 from. http://www.bus.lsu.edu/accounting/faculty/lcrumbley/edutain.html

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.

Bangert, A. W. (2008). The development and validation of the student evaluation of online teaching effectiveness. *Computers in the Schools*, *25*(1), 35–47.

Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching. *Practical Assessment, Research, and Evaluation*, *12*(6). Retrieved November 30, 2010 from http://pareonline.net/pdf/v12n6.pdf

Dziuban, C. D. (2011, May). Technology-enhanced learning: Opportunities and challenges. *Presented at the Faculty Conference on Teaching Effectiveness, Miami, FL.*

Dziuban, C. D., Hartman, J. L., Cavanagh, T., & Moskal, P. D. (2010, November). Targets of opportunity: Using the localness concept to reframe educational opportunity. *Presented at the 16th Annual Sloan Consortium International Conference on Online Learning, Orlando, FL.*

Dziuban, C. D., Moskal, P. D., Brophy-Ellison, J., & Shea, P. (2007). Technology-enhanced education and millennial students in higher education. *Metropolitan Universities*, *18*(3).

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361.

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, *43*(4), 483–501.

Felder, R. (1992). What do they know anyway? *Chemical Engineering Education*, *26*(3), 134–135.

Feldman, K. A. (1976). The superior college teacher from the student's view. *Research in Higher Education*, *5*(243–288).

Ferguson, J., & Tryjankowski, A. M. (2009). Online versus face-to-face learning: Looking at modes of instruction in Master's-leveled courses. *Journal of Further and Higher Education*, *33*(3), 219–228.

Figlio, D. N., Rush, M., & Yin, L. (2010). Is it live or is it Internet? Experimental estimates of the effects of online instruction on student learning. *NBER working paper No. 16089)*. Retrieved November 29, 2010 from The National Bureau of Economic Research Web site: http://www.nber.org/papers/w16089

Greenwald, A. G., & Gilmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1209–1217.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18* 277-269.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19*, 149–161.

Harmon, K. (2004). You are here: Personal geographies and other maps of the imagination. New York: Princeton Architectural Press.

Harris, M. L., & Harris, C. W. (1971). A factor analytic interpretation strategy. *Journal of Educational and Psychological Measurement*, *31*(3), 589–606.

Hartman, J. L., Dziuban, C., & Brophy-Ellison, J. (2007). Faculty 2.0. *Educause review*, *September/October 2007*.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70.

Hoyt, D. P. (1977). *Development of the IDEA system* (Tech. Rep. No. 1). Manhatten, NY: Kansas State University, Center for Faculty Evaluation and Development.

Jaggers, S. S., & Bailey, T. (2010). Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis. Retrieved November 29, 2010 from Teachers College, Columbia University, Community College Research Center Web site. http://ccrc.tc.columbia.edu/

Kaiser, H. F. (1976). Review of Lawley, D. N., & Maxwell, A. E. Factor Analysis as a Statistical Method. *Educational and Psychological Measurement*, *36*, 586–589.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Journal of Educational and Psychological Measurement*, *34*(1), 111–117.

Kim, C., Damewood, E., & Hodge, N. (2000). Professor attitude: Its effect on teaching evaluations. *Journal of Management Education*, *24*(4), 458–473.

Kolitch, E., & Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education*, *24*(1), 27–43.

Linn, R. L., Centra, J. A., & Tucker, L. (1975). Between, within, and total group factor analyses of student ratings of instruction. *Multivariate behavioral research*, *10*(3), 277–288.

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

Mayes, T. (2001). Learning technology and learning relationships. In J. Stephenson (Ed.), *Teaching and Learning Online* (pp. 16). Sterling, VA: Stylus Publishing.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, *52* (11), 1218–1225.

Moore, J. C. (2005). The Sloan consortium quality framework and the five pillars. Newburyport, Mass: The Sloan Consortium.

Mulaik, S. (1972). The Foundations of Factor Analysis. New York: McGraw-Hill.

RateMyProfessors (2011). Retrieved January 3, 2010 from. http://ratemyprofessors.com/index

Russell, T. L. (2001). The no significant difference phenomenon: A comparative research annotated bibliography on technology for distance education (5th ed.). Montgomery, AL: ARELLO Education and Technology Division.

Shelvin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, *25*(4), 397–405.

Shirky, C. (2008). Here comes everybody: The power of organizing without organizations. New York: Penguin Books.

Stewart, I., Hong, E., & Strudler, N. (2004). Development and validation of an instrument for student evaluation of the quality of Web-based instruction. *The American Journal of Distance Education*, *18*(3), 131–150.

Thurstone, L. L. (1935). The vectors of the mind. Chicago: University of Chicago Press.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* DC: Washington.

Wang, M. C., Dziuban, C. D., Cook, I. J., & Moskal, P. D. (2009). Dr. Fox Rocks: Using data-mining techniques to examine student ratings of instruction. In M. C. Shelley, L. D. Yore, & B. Hand (Eds.), *Quality research in literacy and science education: International perspectives and gold standards* (pp. 383–398). Dordrecht, The Netherlands: Springer.

YouTube (2010). Retrieved November 29, 2010 from. http://www.youtube.com/