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