Georgia State University
ScholarWorks @ Georgia State University

Managerial Sciences Faculty Publications | Department of Managerial Sciences

2012

An Empirical Investigation of Student Evaluations of Instruction - The Relative Importance of Factors

Satish Nargundkar
Georgia State University

Milind Shrikhande
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/managerialsci_facpub

Part of the Educational Assessment, Evaluation, and Research Commons, and the Management Sciences and Quantitative Methods Commons

Recommended Citation
Nargundkar, S. & Shrikhande, M. (2012). An Empirical Investigation of Student Evaluations of Instruction—The Relative Importance of Factors. *Decision Sciences Journal of Innovative Education 10*(1), 117-135. doi: 10.1111/j.1540-4609.2011.00328.x

This Article is brought to you for free and open access by the Department of Managerial Sciences at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Managerial Sciences Faculty Publications by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

An Empirical Investigation of Student Evaluations of Instruction – The Relative Importance of Factors

Satish Nargundkar *
Department of Managerial Sciences
Robinson College of Business
Georgia State University
Atlanta GA 30302
678-644-6838
snargundkar@gsu.edu

Milind Shrikhande
Department of Finance
Robinson College of Business
Georgia State University
Atlanta GA 30302
404-406-8556
mshrikhande@gsu.edu

* Communicating Author

Acknowledgement: We thank Harry Dangel for his detailed comments and encouragement. We also thank Yezdi Bhada and Harvey Brightman for providing us with details of their 1989 study that we have replicated here.

An Empirical Investigation of Student Evaluations of Instruction – The Relative Importance of Factors

Abstract

We analyzed over 100,000 student evaluations of instruction over four years in the college of business at a major public university. We found that the original instrument that was validated about 20 years ago is still valid, with factor analysis showing that the six underlying dimensions used in the instrument remained relatively intact. Also, we found that the relative importance of those six factors in the overall assessment of instruction changed over the past two decades, reflecting changes in the expectations of the current millennial generation of students. The results were consistent across four subgroups studied – Undergraduate Core, Undergraduate Non-Core, Graduate Core and Graduate Non-Core classes, with minor differences. *Student Motivation* (the instructor's ability to motivate students) and *grading/assignments* (fairness and objectivity of grading practices) have superseded *presentation ability* in relative importance as indicators of overall teaching effectiveness. Our study has implications for teachers in terms of the appropriate areas to focus on for improving their teaching practices.

1. Introduction

Student evaluations of instruction (SEI) are considered essential for providing feedback to instructors, while also providing administrators with a metric for measuring the effectiveness of instruction for faculty related personnel decisions. From the students' perspective, such feedback is useful if it improves the quality of instruction along dimensions that the students perceive to be most relevant to effective learning. It also helps students make choices about the instructors they want (Marsh & Dunkin, 1992) assuming they are publicly available.

It is important to address the issue of how valid the instruments are in general, and specifically, to test whether the validity of an instrument is maintained over time. There is a perception among some instructors and administrators that the student evaluation process is a ritual with no real implementable value (Abrami et al, 1996). This is typically because instruments in many institutions are developed internally with little research to support the validity of the constructs involved, and little external review or examination.

Marsh (1987) examined a few instruments that did receive some kind of scrutiny. Several studies (Marsh, 1982; Arubayi, 1987, Marsh & Bailey, 1993) provide evidence regarding the reliability of SEIs. For example, the inter-rater reliability of SEIs has been shown to be high, and student evaluations of multiple courses taught by the same teacher are highly correlated, while there is little correlation between student ratings of different instructors, even when rated by the same students. In other words, SEIs do capture the differences between teachers, rather than the course content. Similarly, several studies (Marsh and Hocevar, 1991; Marsh and Roche, 1997; d'Apollonia & Abrami, 1997) have discussed the validity and the multidimensional nature of the SEIs. Some have argued that there is only one global dimension for overall effectiveness of teaching instead of multiple factors. While some extraneous factors (Husbands & Fosh, 1993) such as initial liking for the course and expected grades, among others may contribute to some bias in SEIs, Richardson (2005) found from a survey of the literature on SEIs that such biases are small, and do not impact the overall validity of a well designed instrument.

Marsh (1982) describes an instrument widely used in pedagogical research called the Students' Evaluations of Educational Quality (SEEQ). This instrument has 35 statements broken down into nine factors, namely, *learning/value, enthusiasm, organization, group interaction, individual rapport, breadth of coverage, examinations /grading, assignments,* and *workload/difficulty*. Centra (1993) and Braskamp and Ory (1994) identified six factors commonly found in student rating forms – *course organization and planning, clarity/ communications skills, teacher-student interaction/rapport, course difficulty/workload , grading/examinations* and *student self-rated learning*. The Idea Center (http://www.theideacenter.org/) is another source of well known SEI instruments. It provides customized student evaluation instruments to various universities and has conducted studies on reliability and validity of instruments.

Cohen (1981) examined the relationship between overall instructor ratings and student performance, using data from 67 different multisection courses. He found an overall correlation of +0.43 between the two. Looking into specific dimensions of the SEIs, he concluded that certain dimensions had a greater correlation on student performance than others. Specifically, he mentioned "Skill" and "Structure" as two dimensions that had a greater impact on learning. In

other words, if instructors were well organized, used time wisely and were skilled in presenting the material, their students learned better than the students of other instructors. Cohen (1981) also said that other aspects of teaching, like classroom interaction with students, while positively related, were not very strong, and course difficulty was not related at all to student learning. The greater importance of organization and skill was speculated to be true perhaps for information oriented core classes only, which made up the bulk of the sample. Feldman (1989) extended Cohen's (1981) study, using 17 dimensions related to teaching to find out which ones had the greatest impact on student achievement. He found that the dimensions of teaching that were most highly correlated with student performance were *Course Organization*, *Presentation Clarity*, *Perceived Outcome of Teaching* (Relevance), *Stimulation of Interest* (motivating the study of the subject), followed by other dimensions like *Encouragement Of Classroom Discussion*, *Availability And Helpfulness*. It therefore makes sense that for SEIs to be useful, they must provide feedback to instructors on these dimensions.

In our study, we examine an instrument in use at the Robinson College of Business at Georgia State University, which has 33 items broken down into six factors (see Appendix A for the items used in the instrument, organized according to the six factors). This instrument is a modified version of an instrument validated and used at UC-Berkeley (Brightman et al 1993). Brightman (2005) argues that in order to aid in the improvement of teaching, an instrument must not only be valid, but must provide comparative data to aid interpretation. Peterson et al (2008) conducted one such study recently within one department at a business school. Further, institutional processes must be in place to help faculty diagnose teaching quality and make appropriate improvements.

Brightman et al's (1993) innovation consisted of providing comparative percentile data on the six factors underlying the thirty-three question items rather than merely reporting comparative data on individual question items. They also used four categories to norm the data – undergraduate core, undergraduate non-core, graduate core and graduate non-core. They compared instructor ratings separately for each of these categories. The benefit of showing faculty the scores on each of the six factors was that it helped them focus attention on the broader critical dimensions of their teaching effectiveness. It helped faculty members in diagnosing areas

where they might put forth effort to improve their teaching effectiveness. Brightman et al (1989) authored an internal document at Georgia State University's Robinson College of Business that analyzed the relative importance of the six factors in predicting the overall effectiveness score for the instructor. They found that for the SEI they studied the relative importance of the six factors in determining the overall effectiveness of an instructor was in the following order – 1. *Organization/clarity 2. Presentation ability, 3. Grading /assignments, 4. Student motivation, 5. Student interaction,* and *6.Intellectual / scholarly ability*. Further, they found that these six dimensions together explained over 70% of the variation in overall teaching effectiveness ratings of instructors. The faculty at the College of Business indicated overwhelming support for the new instrument. The implication of Brightman et al's (1989) study and Feldman's (1989) study taken together is that some dimensions of teaching, like *Organization/Clarity* and *Presentation Ability* are not only highly correlated with overall instructor ratings, but also with student learning.

One question that needs to be answered is whether the validity of the instrument still holds up two decades after it was originally implemented. Equally important is ascertaining whether the relative importance of the six factors in predicting the overall effectiveness score for the instructor changes over time. The literature is very vocal about the current generation of students, dubbed the millennial generation; how they are different from their predecessors in terms of their relationships with parents, expectations from college and careers, and other characteristics (Meister & Willyerd, 2010; Koc, 2008; Howe & Strauss, 2003). Some of the key characteristics discussed in the literature suggest that this generation of students have led a sheltered life, are team-oriented, believe that they are special, and are confident in their ability to succeed. They want to connect to a larger purpose, and look to adults to provide guidance and mentoring. This suggests that their perception of good teaching may also be different from that of students a generation ago. For instance, they may value group work, interaction, and consider the teachers' ability to motivate them more important than oratorical skills and presentation ability. Also, the use of technology in the classroom has significantly changed over the past two decades, and may have an impact on what students consider more important. Are classroom interaction and presentation skills just as important to overall teaching effectiveness when all course materials are available on the internet?

About two decades later, we are setting out to answer two questions regarding the instrument. First, are the factors still valid? In other words, do the individual question items still belong to the factors as defined in the Brightman et al (1993) study? To study this, we use a large sample of student responses on these thirty-three question items from ten different departments at the Robinson College of Business at Georgia State University during four consecutive academic years, 2005 through 2009. Second, we examine whether the relative importance of these factors in explaining the overall effectiveness rating of the instructor has changed along with the generation. We therefore replicate the analysis conducted by Brightman et al (1989). Given that initial liking of the course may have something to do with student expectations (Marsh, 1987) from the instructor and the course, the analysis of the relative importance of the factors needs to be conducted separately for both core and non-core classes. Core classes are mandatory, and hence will likely have a higher percentage of students with low initial liking, while non-core classes (electives) are selected by students based on their interest, and should represent a high initial liking of the course. Brightman et al (1993) suggest four norming groups that are meaningful – graduate core, graduate non-core, undergraduate core and undergraduate non-core. We therefore analyze the data for the same four groups.

In the next section, we discuss the data collected and the analysis performed. In section 3, we present the findings from our study. In section 4, we interpret the results and discuss the implications for teachers. Finally in section 5, we discuss limitations of the study and future research.

2. Data Collection and Methodology

2.1 Sample Data

Our large sample from four consecutive academic years spans the ten major departments at the College of Business, as well as an eleventh category called Dean's Office Business Administration, which includes classes like "Master's Orientation", that do not fit into any department. The student responses in the sample are anonymous but information is available about course number and semester. The courses were classified into four groups - Undergraduate Core, Undergraduate Non-core, Graduate Core and Graduate Non-core, and the

responses were categorized accordingly. **Table 1** shows the number of responses in our sample aggregated across four academic years from 2005 to 2009.

	UN	UC	GN	GC	Total
Accounting	8,869	8,972	3,671	2,215	**23,727**
Computer Information Systems	1,551	3,513	1,571	865	**7,500**
Dean's Office – Business Admin	39	***	248	2,334	**2,621**
Finance	3,567	3,382	2,403	1,053	**10,405**
Health Admin	84	***	694	***	**778**
Hospitality Admin	5,037	***	1	***	**5,038**
International Business	480	3,846	1,231	***	**5,557**
Managerial Sciences	5,829	10,405	4,442	3,160	**23,836**
Marketing	4,787	6,992	1,836	2,546	**16,161**
Real Estate	2,818	***	881	***	**3,699**
Risk Management & Insurance	3,122	4,138	1,399	1,362	**10,021**
Total Responses	**36,183**	**41,248**	**18,377**	**13,535**	**109,343**
Total Enrollment	**57,971**	**47,699**	**26,909**	**21,945**	**154,524**
Response Rate	**62.42%**	**86.48%**	**68.29%**	**61.68%**	**70.76%**

Table 1: Responses by Department and Segment

*** These departments did not have any classes in this segment.

UN = Undergraduate Non-Core
UC = Undergraduate Core
GN = Graduate Non-Core
GC = Graduate Core

Note that the enrollment in non-core classes exceeds the enrollment in core classes (e.g. 57,971 vs 47,699 for Undergraduate), simply reflecting the fact that a greater number of non-core classes are offered compared to core classes at both the graduate and undergraduate levels. Richardson (2005) surveyed the literature on student evaluation instruments, and indicates that response rates of around 60% are common and that a 70% response rate would be considered good. The overall response rate across all categories in our study is just over 70%, indicating a good overall response rate. The response rates for the four subgroups as shown in the table above vary from about 62% to about 86% showing agreement with Richardson's (2005) study. The response rate for Undergraduate Core classes is especially good at 86.48%. It is interesting to note that the response rate at the undergraduate level is lower for the non-core classes, while at the graduate level, it is lower for the core classes. These evaluations are filled out online, and students are given an incentive – when they are done filling out the SEI, they get to see their course grades. They do have the option to decline, in which case they can see the grades, but

cannot come back to complete the SEI. There is the possibility of selection bias, since students that respond are different from those who do not in terms of their study habits and academic achievement (Watkins & Hattie, 1985; Nielsen et al, 1978). Such a bias may be unavoidable, though its effect on our study is likely to be small, given that we are comparing responses over time, thus studying the changes among the respondents over time.

2.2 Methodology

We used factor analysis to revalidate the factors underlying the thirty-three question items. The factor analysis approach can be briefly described as follows. The SEI instrument used is a modified version of the instrument validated at UC Berkeley (Brightman, et al. 1993). It was designed to represent six factors measured by thirty three items in all. The factor analysis was performed using maximum likelihood estimates with six factors pre-specified (Segars et al, 1993). Varimax rotation was performed to separate the factors clearly (Kaiser, 1958). Items were associated with the factor with which they had the highest loading. These item combinations were studied to see if they matched our expectation based on the Brightman et al (1993) study. This analysis was performed separately for each of the four groups, for data across four academic years between 2005 and 2009, giving us a large sample of over 100,000 observations.

The second step in the analysis was to perform stepwise regression (Hocking, 1976) with the score on question 34 (a global evaluation of the instructor) as the dependent variable and the six factor scores as the independent variables. The intent was to discover whether the factor scores were related to the overall score, and to discover the relative importance of the factors. It also helped to determine the cumulative impact of all six factors in determining the overall effectiveness score of the instructor. Thompson (1995) has highlighted some problems with the stepwise procedure in terms of its being able to identify the best combination of variables and potentially inflated significance. However, for this study, the procedure makes sense for several reasons. First, the factors have already been identified. The goal is not to find a subset of variables that best describe the model, but to rank order them. Second, we use a very large sample size, which reduces the impact of problems with degrees of freedom leading to any spurious significance. Finally, we are replicating a study (Brightman, et al, 1989) that used the

same method to rank order the factors initially, and a comparison using the same method will show us if anything else has changed in the nature of the relationships over the two decades separating the studies.

3. Results

We first present the factor analysis to examine the item loadings on each factor and then the stepwise regression analysis to study the relative importance of the factors. Table 2 below presents the factor analysis results. It shows how the individual question items were categorized into different factors in the original studies (Brightman et al, 1993, 1989) and how they have changed over time. We do this comparison for each of the four groups (Undergraduate Core and Non-core, Graduate Core and Non-core).

	Factor 1 **Prsntation**	**Factor 2** **Org/Clarty**	**Factor 3** **Grdng/Asn**	**Factor 4** **Intel/Sch**	**Factor 5** **Interact**	**Factor 6** **Motivtn**
Original (1993)	19,20,22, 23,24,25, 26,27	5,6,11,12, 13,14,15, 18	1,2,3,4, 31,32,33	7,8,9,10	16,17,21	28,29,30
Undergraduate Core n=41,248	-19, -20	-5, -14, -15	+5, +14, +15, +19		+20	
Undergraduate Non-core n=36,183	-20	-5, -14, -15	+5, +14, +15		+20	
Graduate Core n=13,535	-20	-5, -14, -15	+5, +14, +15	+17	-17,+20	
Graduate Non-core n=18,377	-20	-5, -14, -15	+5, +14, +15		+20	

Table 2: Factor Analysis of SEI items by Segment

The numbers in the table show the items that did not match the original classification. A plus sign indicates that an item loaded heavily on the given factor, and a minus sign means that it loaded lower than expected. For instance, question items 5, 14, and 15 were originally supposed

to be part of the Organization/Clarity factor, but in our study they consistently loaded more heavily on the Grading/Assignments factor. As the table shows, items 5, 14, 15 and 20 were consistently classified differently than in the original study. Items 16 and 17 had loadings that were generally diffused and did not clearly load on any factor. [1]

A sample output of the factor analysis (for Undergraduate Non-Core) is shown in Appendix B. While there was some movement of items across factors compared to the original study, such as the ones mentioned above, most of the items loaded heavily on the factors that they were originally part of, as expected.

We see from Table 2 that the re-categorization of individual question items by factors is uniformly observed across the four categories we studied. This indicates, as expected, that the loading of items on a given factor is independent of whether it is a core course or a non-core one, or whether it is at the graduate or undergraduate level. Overall, we find that the six factors as outlined in the original study have remained largely intact. While some question items loaded differently across factors, the differences were judged by us to be marginal in nature.

We now present below the results of the stepwise regression performed to determine the relative importance of the factors. Table 3a shows the re-ordering of the relative importance of factors for Undergraduate Core and Non-Core classes compared to the original study[2], while Table 3b does the same for Graduate Core and Non-Core classes.

[1] The analysis was also performed separately for each year from 2005 to 2009 for both core and non-core classes, and the results were consistent with those presented in Table 2, for each of the four years.

[2] Brightman et al (1989) report that the differences across the four groups in their study were minimal, and the relative importance of factors was the same. We were only able to obtain one set of cumulative R-squares from their study, shown in tables 3a and 3b.

Order of Importance of Factors		New order of importance			
Original Study	**Cum. R-Square**	**Undergraduate Core**	**Cum. R-Square**	**Undergraduate Non-core**	**Cum. R-Square**
Organization/Clarity	25.90	Organization/Clarity	23.26	Organization/Clarity	25.52
Presentation Ability	45.50	**Student Motivation**	39.95	Grading/Assignments	41.37
Grading/Assignments	53.90	Grading/Assignments	51.45	**Student Motivation**	54.26
Student Motivation	59.90	Student Interaction	58.77	Student Interaction	65.33
Student Interaction	65.80	**Presentation Ability**	64.47	**Presentation Ability**	73.54
Intellectual/Scholarly	70.50	Intellectual/Scholarly	68.52	Intellectual/Scholarly	75.04
$p<0.001$ for all factors		n=41,248		n=36,183	

Table 3a: Stepwise Regression Results, Undergraduate Core and Non-Core, 2005-2009

Brightman et al (1989) found that the two factors ranked as the most important in explaining the overall effectiveness of the instructor were *Organization / Clarity* and *Presentation Ability*. Our study reveals that the two most important factors now seem to be *Organization / Clarity* and *Student Motivation* for undergraduate core classes, while *Grading/Assignments* takes second place for undergraduate non-core classes. *Presentation Ability* dropped down to the fifth rank among the six factors in both groups.

Order of Importance of Factors		New order of importance			
Original Study	**Cum. R-Square**	**Graduate Core**	**Cum. R-Square**	**Graduate Non-core**	**Cum. R-Square**
Organization/Clarity	25.90	Grading/Assignments	24.33	Organization/Clarity	24.95
Presentation Ability	45.50	Organization/Clarity	44.07	Grading/Assignments	44.98
Grading/Assignments	53.90	**Student Motivation**	56.68	**Student Motivation**	58.40
Student Motivation	59.90	**Presentation Ability**	67.16	**Presentation Ability**	67.11
Student Interaction	65.80	Intellectual/Scholarly	73.31	Student Interaction	71.93
Intellectual/Scholarly	70.50	Student Interaction	78.52	Intellectual/Scholarly	75.54
$p<0.001$ for all factors		n=13,535		n=18,377	

Table 3b: Stepwise Regression Results, Graduate Core and Non-Core, 2005-2009

For graduate classes, our study reveals that *Presentation Ability* dropped down to the fourth rank among the six factors in both groups (core and non-core). The two most important factors are

Organization / Clarity and *Grading/Assignments* (though their order is interchanged between the two groups) followed by *Student Motivation*.

4. Discussion / Reflections

As discussed before, the millennial generation is regarded as being different from the previous generation in terms of their expectations in school and in the workplace. Some of these differences may have translated into the results we saw regarding the factors that they consider most important to effective instruction. Students of the millennial generation have been characterized as feeling that they are special, having lived a more sheltered life, and trusting institutions and adults to guide them in their career and life choices. In other words, they rely more on adults to mentor and motivate them. Our results seem to reflect this situation in the sense that student motivation is now a more important component of overall teaching effectiveness than before. Some of this change that is reflected in student expectations is perhaps attributable to the changes in teaching practices and technology over the years.

The significant decline in the relative importance over time of *Presentation Ability* and the increased importance of *student motivation* and *grading/assignments* can be interpreted as a shift from a teacher-centered classroom to a student-centered classroom. *Presentation Ability* has been ranked fifth instead of second for undergraduate students and fourth for graduate students. *Student Motivation* has climbed to second place for undergraduate core classes, and to third for all other groups. *Grading/assignments* remains unchanged at number three for undergraduate core classes, and has increased in importance for the other groups. The implication of these significant changes is that overall teaching effectiveness is determined more by how well the instructor motivates students to learn the material and how fair, consistent and structured the grading and assignments are, rather than simply presentation ability.

Ryan &Deci (2000) examine the classic definitions of intrinsic and extrinsic motivation and discuss the various levels of autonomy that can exist with extrinsic motivation. They define intrinsic motivation as the desire to do something because it is inherently enjoyable, and extrinsic motivation as the desire to do something for a separable outcome, and argue that there are various types of extrinsic motivation in terms of the autonomy or self-determination that they

allow the student. One extreme, with no autonomy, is that of extrinsic rewards and punishments such as earning a grade for the course. The student response in such a case is characterized by reluctance. A student may however want to do work for the sake of approval from oneself or others, which is a mix of external and internal motivation. At the other end of the spectrum, a student may consciously value the goals set by an instructor, and eventually even internalize them. In other words, instructors do have a role in motivating students, and it is not merely the traditional idea of external rewards. Rather, an instructor can serve as a catalyst to awaken the intrinsic motivation of a student, by convincing them of the value of the goals set for a class. Deci & Ryan (1985) discuss the *Self-Determination Theory* (SDT), and suggest that instructors foster the internalization and integration of values in order to motivate students.

The process of convincing students about the value of goals begins with setting clear objectives for the class, and following the plan for the course that is laid out in the syllabus. The next step is to ensure that assignments and tests are related to these objectives. Finally, one must ensure that the grading/evaluation are fair, objective, and consistent. All of these elements are reflected in the factor called *grading/assignments*. The increased importance of this factor in our study is an indicator that students today value these elements of teaching more than ever. The g*rading/assignments* factor has increased in importance relative to the original study for the undergraduate non-core group and for both the core and non-core graduate groups. However, for the undergraduate core group, student motivation is more important than grading/assignments, since they are the least likely group to be self-motivated.

It is important to note that it is not the leniency in grading that matters. Marsh & Roche (2000) summarize the literature debunking the myth that higher SEI ratings can be obtained by reducing the workload and being more lenient in grading. Centra (2003) also conducted an empirical study where he analyzed data from over 50,000 college courses and found little reason to believe that inflating grades produces better SEI ratings. Across several studies, a small positive correlation (about 0.20) has been found between a student's expected grade and the SEI rating of the instructor. However, Centra (2003) and Marsh & Roche (2000) conclude that this correlation has several possible explanations which do not involve bias in the grading process.

Skinner & Belmont (1993) studied motivation of students in elementary school (grades 3-5) and teachers who provided students with autonomy and optimal structure had students that were motivated throughout the school year. Lin, et al (2003) found that students that had a high intrinsic motivation and a medium level of extrinsic motivation tended to achieve the highest grades, suggesting that extrinsic motivation, if not taken to an extreme, does have a role to play in performance. Dan Pink (2009) in his talk on TED.com, summarizes findings from several studies by economists and social scientists, mainly Ariely, et al (2005), and highlights three things that motivate people in the workplace – autonomy, mastery, and purpose. For menial tasks, money is a good motivator. However, for tasks requiring cognitive skills, money only matters up to a point and too much emphasis on it can actually demotivate. After a certain level of earning, people are motivated by the need to learn and master things, to be self-directed, and work for a purpose they believe in. The implication of all this in the college classroom is that first, instructors must be organized and set a structure for the course with objectives that students find relevant and challenging, to give them a higher purpose than simply memorizing certain content. Second, instructors must establish fair evaluation systems, and then provide students with sufficient autonomy to engage them with the material. Our results, as reflected in the increased importance of *student motivation* and *grading/assignments* suggest that the instructor's ability to show students the relevance of the material to their life or to a higher purpose is important to them.

Student Evaluation of Instruction (SEI) instruments are now well established in various parts of the world, especially in the United States and Australia (Richardson, 2005). This proliferation of SEIs over the last generation has contributed to a greater attention to teaching methods by instructors, administrators and students alike. There has been a growing emphasis on active learning methods, leading to more student-centered approaches to instruction. The ability to motivate students is therefore more indicative of teaching effectiveness today than mere presentation ability.

These findings further encouraged us to look at the difference in the importance of factors relating to effectiveness of instruction at the core versus the non-core levels. As one would expect, undergraduate students choose non-core classes based on their interest, and are therefore

a little less concerned about external motivation to study them. The results bear this out by showing that overall effectiveness ratings are slightly less dependent on student motivation (ranked third) for non-core classes than for core classes. Further, one would expect a wider variety of students in core classes than in non-core classes. By definition, non-core classes have students in them that are a more homogenous group in terms of their interest in the subject. One would thus expect that a predictive model for students in the non-core classes would have lower variation or error than for their counterparts in the core classes. Consistent with the expectation, the regression results for undergraduate students show us that about 68.5% of the variation in overall scores can be explained by the six factors for core classes and a significantly higher 75% for non-core classes. For graduate students also, student motivation was found to be more important than presentation ability. It was ranked third for both core and non-core classes, since, like the students in the undergraduate non-core segment, one expects graduate students to be more self-motivated overall. Contrary to expectation, the cumulative R-square for graduate core classes at 78.5% was higher than for graduate non-core at 75.5%. All the cumulative R-squares, ranging from 68.5% to 78.5% are high enough to indicate that the six factors used in the instrument are all important to the overall effectiveness of instruction.

While we see that the advent of the millennial generation did impact the relative importance of factors, their effect on the basic validity of factors themselves is at best marginal. The validity of the instrument demands that the items purporting to measure certain underlying factors still be correlated with each other. The generational change should not have an impact on the grouping of items into the underlying six factors. Our results show that a few of the items loaded more heavily on a factor other than the one expected. This could be partially due to an overlap between the constructs. Paswan and Young (2002) have studied the nomological relationships among five constructs in another such instrument, and found significant influences of some factors on the others. In other words, the underlying constructs are not truly independent of each other. However, as we saw with the regression analysis, the slightly altered factors still explained about the same amount of variation in the overall effectiveness score as they did almost twenty years ago.

While an instrument may be validated in a research study, improper implementation of the instrument can undo its validity. One of the concerns some faculty members have with student evaluations is that students may not pay enough attention to the questions, simply rating the instructor overall and then marking the same number all the way down for each question, to simply 'get it over with'. Instructors report anecdotal data regarding students completing an SEI administered in-class at the end of a semester in a minute or two, which would be impossible if the student took the time to read and answer 35 questions. If that were the case, however, the items would all be correlated to each other, and would not resolve themselves into factors very well. Our results show that this was not the case, and the items did combine into factors as expected. The anecdotal evidence against this may represent a very small fraction of students, or those students who finish in very little time may in fact be leaving the survey blank. Also, our sample over the past four years represents SEIs completed online, which students can complete when they have the time, thereby potentially increasing the likelihood that they will complete it appropriately.

In summary, we believe that our study makes the following contributions towards improving teaching that should be generalizable across disciplines and universities. First, the study shows that a well- designed and validated SEI holds the promise of remaining robust over many years. The instrument must enable not merely a summative assessment of past teaching effectiveness, but provide guidance to the instructor on changes and innovations that they could make to improve their effectiveness. The individual items in the SEI must be combined into factors to provide feedback on dimensions critical to effective teaching, as shown earlier in the Brightman et al (1993) study. Also, based on Feldman's (1989) study, some of the same factors that are important in predicting overall instructor effectiveness are also correlated with student performance. Further, our study shows that it is valuable to know the relative of importance of these factors as well as the changes in relative importance over time due to generational shifts. This will enable instructors to decide on the aspects of their teaching to focus on for improving their effectiveness as well as student learning.

5. Future Research

Using SEIs as a way to improve teaching and to evaluate teaching performance requires that several key elements be put in place. The first requirement is the existence of a valid and reliable instrument. Second, there should be a mechanism to fairly compare ratings across instructors, and finally, an institutional process in place for faculty development. This study confirms the validity of the instrument, and provides insights into the changing nature of the relative importance of the factors measured by the SEI. There is still the issue of comparing scores across instructors. Currently, the comparisons are made in four separate segments - Core vs. Non-Core classes, and Undergraduate vs. Graduate classes. In other words, all instructors teaching undergraduate core classes are compared with one another. Are there other factors that are relevant in making these comparisons and need to be controlled for? Some questions to consider in the future may be – does time of the day that the class meets matter? Are ratings different by location (downtown vs. other campuses) or by department? One can also study the differences in ratings based on rank of the instructor, gender of the instructor (gender bias), gender of the student to see if there are any systematic ratings differences. The issue of grade inflation is another avenue for research. Does the grade distribution correlate positively with instructor ratings? The study conducted by Peterson et al (2008) at Montclair State University's School of Business is worth replicating at another institution to see if the effects of extraneous factors on SEI ratings are consistent across institutions.

Most importantly, since SEIs are supposed to help teachers improve, is there evidence of improvement over time? We plan to study the ratings over time for all instructors over the same four year period to see if there were improvements.

In the context of motivation of students to perform better, Pintrich (2003) surveyed the literature and suggests several questions that have been studied and can still be asked, including (1) what motivates students in classrooms? (2) Do students know what motivates them? (3) What is the relationship between motivation and cognition? (4) How does motivation change and develop? These questions and others like them merit further research.

References

Abrami, P. C., d'Apollonia, S. & Rosenfield, S. (1996) The dimensionality of student ratings of instruction: what we know and what we do not, in: J. C. Smart (Ed.) *Higher education: handbook of theory and research,* volume 11 (New York, Agathon Press).

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2005) Large Stakes and Big Mistakes, Federal Reserve Bank of Boston, Working Papers, 5-11, July.

Arubayi, E. A. (1987) Improvement of instruction and teacher effectiveness: are student ratings reliable and valid?, *Higher Education,* 16, 267–278.

Brascamp, L.A. and Ory, J.C. (1994). *Assessing Faculty Work. Enhancing Individual and Institutional Performance*. San Francisco: Josey Bass.

Brightman, H.J., Bhada, Y., Elliott, M., & Vandenberg, R. (1989) An Empirical Study to Examine the Reliability and Validity of a Student Evaluation of Instructor Instrument, GSU College of Business Administration Internal Working Document, prepared by the Faculty Development Committee (FDC).

Brightman, H., Elliott, M., Bhada, Y., (1993) Increasing the Effectiveness of Student Evaluation of Instructor Data through a Factor Score Comparative Report, *Decision Sciences*, Jan/Feb, p. 192-199

Brightman, H., (2005) Mentoring Faculty to Improve Teaching and Student Learning, *Decision Sciences Journal of Innovative Education*, Volume 3, No. 2, pp. 191-203.

Centra, J. A. (1993) *Reflective Faculty Evaluation*. San Francisco, Jossey-Bass.

Centra, John A. (2003) Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work? *Research in Higher Education,* Vol. 44, No. 5, October.

Chen, Y. (2003) Student evaluation of teaching effectiveness: An assessment of student perception and motivation, *Assessment and Evaluation in Higher Education* (Routledge).

Cohen, P.A. (1981) Student Ratings of Instruction and Student Achievement, *Review of Educational Research*, 51(3), 281-309.

D'Appollonia, Sylvia, & Abrami, Philip C. (1997) Navigating Student Ratings of Instruction. *American Psychologist* 52 (11): 1198-1208.

Feldman, K. (1989) The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement, *Research in Higher Education*, 30(6) 1989. pp. 583-645.

Kaiser, Henry F, (1958) The Varimax Criterion for Analytic Rotation in Factor Analysis *Psychometrika*.

Hocking, R. R. (1976) The Analysis and Selection of Variables in Linear Regression, *Biometrics*, 32

Howe, N., & Strauss, W. (2003). *Millennials go to college*. United States: American Association of Collegiate Registrars.
Husbands, C. T. & Fosh, P. (1993) Students' evaluation of teaching in higher education: experiences from four European countries and some implications of the practice, *Assessment and Evaluation in Higher Education,* 18, 95–114.

Koc, E. (2008). NACE research: The many faces of the millenial generation. *NACE Journal*, *69*(1), 18.

Marsh, H. W. (1982) SEEQ: a reliable, valid and useful instrument for collecting students' evaluations of university teaching, *British Journal of Educational Psychology,* 52, 77–95.

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

Marsh, H. W. & Bailey, M. (1993) Multidimensional students' evaluations of teaching effectiveness, *Journal of Higher Education,* 64, 1–18.

Marsh, H. W. & Dunkin, M. J. (1992) Students' evaluations of university teaching: a multidimensional perspective, in: J. C. Smart (Ed.) *Higher education: handbook of theory and research, volume 8* (New York, Agathon Press).

Marsh, H. W. & Hocevar, D. (1991) Multidimensional perspective on students' evaluations of teaching effectiveness: the generality of factor structures across academic discipline, instructor level, and course level, *Teaching and Teacher Education,* 7, 9–18.

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, 1187–1197.

Marsh, H. W. & Roche, L. A. (2000) Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders? *Journal of Educational Psychology*, Vol. 92, No. 1. 202-228.

Lin, Yi-Guang, McKeachie, Wilbert J., Kim, Yung Che (2003) College student intrinsic and/or extrinsic motivation and learning", *Learning and Individual differences* Vol. 13, No. 3, pages 251-258.

Meister, Jeanne C., & Willyerd, Karie (2010) Mentoring Millennials. *Harvard Business Review*, May 2010, pp. 68-73.

Paswan, Audhesh, and Joyce A. Young (2002) A Causal Analysis of Process Variables in Student Evaluation of Instructors, *Journal of Marketing Education*, Vol. 24, No. 2, 193-202.

Peterson, Richard L., Berenson, Mark L., Misra, Ram B. and Radosevich, David J. (2008) An Evaluation of Factors Regarding Students' Assessment of Faculty in a Business School, *Decision Sciences Journal of Innovative Education,* Volume 6 Number 2, pp. 375-402.

Pink, Dan, (2009) The Surprising Science of Motivation , Talk onTED.com, http://www.ted.com/talks/lang/eng/dan_pink_on_motivation.html

Pintrich, Paul R., (2003) A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching contexts, *Journal of Educational Psychology*, Vol. 95, No. 4, 667-686.

Richardson, John, T. E., (2005) Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, Vol. 30, No. 4, August 2005, pp. 387-415.

Deci, Edward L. and Ryan, Richard M. (1985) *Intrinsic Motivation and Self-Determination in Human Behavior*, New York, Plenum.

Ryan, Richard M. and Deci, Edward L. (2000) Intrinsic and extrinsic motivations: Classic definitions and new directions, *Contemporary Educational Psychology*, Vol. 25, Issue 1, January 2000, pages 54-67.

Segars, Albert H, Grover, Varun, (1993) Re-examining perceived ease of use and usefulness: A confirmatory factor analysis. *MIS Quarterly*. Minneapolis: Dec Vol. 17, Issue 4; pg. 517, 9 pages.

Skinner, Ellen A. & Belmont, Michael J, (1993) Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year, *Journal of Educational Psychology*, Vol. 85(4), December 93, 571-581.

Thompson, B. (1995) Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial, *Educational and Psychological Measurement.*

Appendix A

Questions on SEI Instrument by Underlying Factor

Grading / Assignments	Organization / Clarity
1. Follows the plan for the course as established in the syllabus 2. Gives assignments related to the goals of this course. 3. Explains the grading system clearly. 4. Is accessible to students out of class. 31. Given nature of exams and assignments, returns them quickly. 32. Gives assignments and exams that are reasonable in length and difficulty. 33. Assigns grades fairly and impartially.	5. Is well prepared. 6. Speaks in a manner that is easy to understand. 11. Explains clearly. 12. Lectures easy to outline or case discussion well organized. 13. Is careful and precise in answering questions. 14. Summarizes major points. 15. States objectives for each class session. 18. Knows if the class is understanding him/her or not.
Presentation Ability	**Student Motivation**
19. Cares about the quality of his/her teaching. 20. Has a genuine interest in students. 22. Is a dynamic and energetic person. 23. Has an interesting style of presentation. 24. Seems to enjoy teaching. 25. Is enthusiastic about his/her subject. 26. Seems to have self-confidence. 27. Varies the speed and tone of his/her voice.	28. Made me work harder than in most other courses. 29. Motivates me to do my best work. 30. Gives examinations requiring creative, original thinking.
Student Interaction	**Intellectual/Scholarly Ability**
16. Encourages class discussion. 17. Invites criticism of own ideas. 21. Relates to students as individuals.	7. Discusses points of view other than his/her own. 8. Contrasts implications of various theories. 9. Discusses recent developments in the field. 10. Presents origins of ideas and concepts.

Global Questions

34. Overall Effectiveness of Instructor
35. Overall Rating of Course Content

Appendix B

Rotated Factor Loadings and Communalities
Undergraduate Non-Core
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Communality
	Grading	Prsntn	Org/Clr	Intrtn	Motvtn	Intel/Sch	
_Q1_	0.707	0.262	0.285	0.181	-0.235	0.191	0.774
_Q2_	0.651	0.319	0.299	0.217	-0.265	0.232	0.787
_Q3_	0.631	0.261	0.296	0.264	-0.247	0.201	0.725
_Q4_	0.558	0.296	0.254	0.339	-0.237	0.230	0.688
_Q5_	**0.656**	0.346	0.365	0.196	-0.234	0.229	0.830
_Q6_	0.439	0.362	0.605	0.205	-0.220	0.252	0.843
_Q7_	0.423	0.305	0.401	0.335	-0.267	0.465	0.832
_Q8_	0.464	0.319	0.370	0.272	-0.292	0.504	0.867
_Q9_	0.452	0.425	0.302	0.213	-0.219	0.510	0.830
_Q10_	0.476	0.367	0.367	0.232	-0.276	0.482	0.860
_Q11_	0.418	0.333	0.664	0.246	-0.242	0.249	0.907
_Q12_	0.526	0.288	0.568	0.256	-0.260	0.195	0.854
_Q13_	0.463	0.325	0.545	0.329	-0.257	0.259	0.858
_Q14_	**0.542**	0.319	0.457	0.303	-0.255	0.263	0.831
_Q15_	**0.559**	0.298	0.371	0.292	-0.277	0.265	0.772
_Q16_	0.389	0.372	0.347	0.402	-0.286	0.323	0.758
_Q17_	0.348	0.282	0.386	0.482	-0.328	0.356	0.816
_Q18_	0.339	0.289	0.571	0.403	-0.343	0.214	0.851
_Q19_	0.467	0.459	0.363	0.451	-0.250	0.223	0.876
_Q20_	0.400	0.481	0.312	**0.584**	-0.231	0.202	0.924
_Q21_	0.370	0.449	0.344	0.567	-0.259	0.209	0.890
_Q22_	0.286	0.600	0.373	0.348	-0.357	0.195	0.868
_Q23_	0.274	0.492	0.448	0.322	-0.440	0.192	0.851
_Q24_	0.391	0.627	0.304	0.375	-0.279	0.201	0.898
_Q25_	0.427	0.678	0.254	0.264	-0.250	0.250	0.901
_Q26_	0.465	0.607	0.274	0.173	-0.268	0.268	0.833
_Q27_	0.310	0.534	0.406	0.238	-0.399	0.232	0.816
_Q28_	0.397	0.314	0.210	0.149	-0.588	0.201	0.709
_Q29_	0.386	0.348	0.367	0.331	-0.564	0.186	0.868
_Q30_	0.452	0.316	0.272	0.286	-0.537	0.224	0.798
_Q31_	0.607	0.314	0.216	0.272	-0.328	0.182	0.729
_Q32_	0.487	0.251	0.341	0.356	-0.366	0.160	0.703
_Q33_	0.540	0.309	0.280	0.390	-0.314	0.197	0.755
Variance	7.4703	5.2111	4.9262	3.5522	3.4335	2.5075	27.1008
% Var	0.226	0.158	0.149	0.108	0.104	0.076	0.821

Note that the factors are presented in order of the variance explained (out of the total of 33). This is dependent on the number of question items on the instrument that correlate to a given factor, and not necessarily an indication of how important the factor is to overall evaluation. The gray highlights are the items that loaded heavily on a given factor. The red numbers show items that showed a weaker than expected correlation, and corresponding boldfaced numbers show the factor with which the item correlated instead. For instance, items 5, 14 and 15 were supposed to be part of Organization/Clarity, but correlated better with Grading/Assignments instead.