Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 1999 Proceedings

International Conference on Information Systems (ICIS)

December 1999

The Role of Media Use and Active Learning in Higher Education: The Development of an Instrument to Determine the Dimensions of Teaching

Mark Serva
University of Delaware

Mark Fuller
Baylor University

THE ROLE OF MEDIA USE AND ACTIVE LEARNING IN HIGHER EDUCATION: THE DEVELOPMENT OF AN INSTRUMENT TO DETERMINE THE DIMENSIONS OF TEACHING

Mark A. Serva
University of Delaware
U.S.A.

Mark A. Fuller
Baylor University
U.S.A.

Abstract

The competition fostered in today's environment is demanding increased accountability of collegiate instructors and re-evaluation of traditional teaching methods. Without reliable and valid instructional measurement systems, it is virtually impossible to benchmark new techniques or identify effective instructors. In addition, these measures need to be comprehensive. This research describes the development of a faculty evaluation instrument that incorporates traditional dimensions of teaching with two new educational concepts: active learning and media use. Structural equation modeling is used to assess reliability and construct and discriminant validity. The results confirm the importance of five historically strong components of teaching, as well as support the importance of media use and active learning within an academic environment.

1. INTRODUCTION

Today's business schools are under increased competitive pressures. Corporations are pushing for not only basic technical knowledge in a field of study, but also skills in teamwork, communication, leadership, and applied decision making (Robbins 1994). At the same time, new education providers are increasing competition for the education dollar. Corporate universities, e.g., Southwest Airlines University for the People (Densford 1997), provide training in many areas formerly relegated to the university classroom. Traditional geographic boundaries to competition are also being crossed, as colleges and universities employ new technology (i.e., the World Wide Web and groupware) to deliver their educational product across city, state, and country lines.

While competitive forces are changing, new regulatory pressures are also occurring. Accrediting bodies (e.g., the AACSB) are asking business schools to devise methods to track performance along their mission and objectives.[1] To further complicate the situation, publications such as *Business Week* have instituted yearly rankings of business schools based on surveys administered to alumni and recruiters. These ranking systems rely heavily on the educational reputation of the institution, rather than just measures of faculty research productivity. In order to excel in these ranking systems, business schools must seek ways of improving the quality of the student's educational experience.

The need to effectively measure teaching quality, therefore, has become obvious. Not only is education the primary mission of any college or university, but the competition fostered in today's environment is demanding increased accountability for what

---

[1]A full explanation of the AACSB's current accreditation policy can be found on the Internet at http://www.aacsb.edu/.

occurs in the classroom. Without good instructional measurement systems, it is virtually impossible to devise meaningful compensation systems that can drive performance improvement in the area of teaching. Finally, these measures need to be comprehensive, covering new advances both in teaching methods and in educational technologies.

2. LITERATURE REVIEW

The dimensions of teaching, and teaching evaluations, have been studied by a variety of researchers. Work by Coffman (1954), Hodgson (1958), and Centra (1973) revealed that the most common dimensions used in measures of instruction included organization, structure, or clarity; teacher-student interaction or rapport; and teaching skill, communication, or lecturing ability. Other dimensions occasionally studied include evaluation of course workload/course difficulty; evaluation of grading/ examinations; evaluation of impact on students (i.e., self-rated accomplishment); and global/overall effectiveness. Wotruba and Write (1975) found similar dimensions in research that summarized 21 studies investigating the qualities of effective teaching, the results of which have been used to develop hundreds of rating forms over the years.

More recent work by Grussing, Valuk and Williams (1994) reaffirmed the results of this earlier research by identifying course organization, teaching ability, grading and feedback, student-instructor interaction, workload and course difficulty, enthusiasm/motivation, and knowledge of subject area as the major dimensions of classroom performance. What is illustrated in a historical view of teaching evaluation research is that a great deal of consistency is apparent across studies done during very different eras. This congruity within the literature suggests that these dimensions can be used as evaluative criteria in the classroom (Seldin 1984; Van Ort 1983).

Despite the consistency of dimensions, however, problems in evaluation still exist. One such problem is the existence of errors that creep into the evaluation process. According to Grussing (1994), a variety of error types exist, including the incorrect interpretation of the meaning of ratings, the impact of other instructor attributes (such as charisma, age, or gender) on the evaluations, a lack of reliability in the ratings, mixed purposes of the ratings (e.g., for evaluation or for merit pay), and instrument content problems.

In the latter case, instrument content problems encompass at least three subtypes of problems (Tagomori and Bishop 1995), including item ambiguity, clarity, and subjectivity. In data collected from over 400 schools of education, Tagomori and Bishop took a sample of 200 SEI (student evaluation of instruction) forms and examined them for flaws. Over 20% of the evaluation questions examined were categorized as ambiguous (e.g., "How clear were the aims, goals, and requirements of the course") and unclear (e.g., "The total experience was very worthwhile") with over 50% being categorized as subjective (e.g., "The class understood the material"). The Tagomori and Bishop study concluded that the validity and reliability of many of the instruments in use in educational institutions should be rigorously questioned.

In addition to the types of errors listed above, educators are also confronted with another issue: "What should we be measuring?" The current educational context is different than it was 10 years ago, and two dimensions of instruction, which have taken on increased importance in the educational literature, are the use of active learning techniques and the application of new media in the learning process.

The concept of active learning is described by Bonwell and Eison (1991) as any activity that "involves students in doing things and thinking about the things they are doing." The principles of active learning consist of two core assumptions: first, learning is by nature an active endeavor; and second, different people learn in different ways (Meyers and Jones 1993). This paper uses the term "active learning" similar to Meyers and Jones to suggest that students have the opportunity to "talk and listen, read, write, and reflect as they approach course content through problem-solving exercises, informal small groups, simulations, case studies, role playing, and other activities, all of which require students to apply what they are learning."

Active learning is not the only way people learn, but many education researchers argue using an active learning style in the classroom will expose students to more opportunities for learning. As Chickering and Gamson (1987) state, "students must talk about what they are learning, write about it, relate it to past experiences, apply it to their daily lives. They must make what they learn as part of themselves." Cross (1987) states that "when students are actively involved in ... learning ... they learn more than when they are passive recipients of instruction." A number of other education researchers have advocated similar stances (Astin 1985; Ericksen 1984).

This model of "active" learning is informed from theory and research on alternatives to the traditional (lecture-based) forms of learning, which pay little attention to "how" students process and retain information. Alternative models of learning, such as constructivist (O'Loughlin 1992) and cooperative (Slavin 1990) models, put a much higher emphasis on the student engagement and participation (i.e., activity) in the learning process.

The use of information technology in education is also increasing. Business schools have started incorporating a variety of new media both to change access to educational materials and to change the educational process itself. Audio and video teleconferencing technologies, presentation software, group decision support systems, simulation and modeling software, and the Internet are all being integrated into educational environments.

Research by Leidner and Jarvenpaa (1993) illustrated that the use of new media to communicate ideas in an academic environment can affect learning. Recent research by Ives and Jarvenpaa (1996) and by Alavi (1994) has examined the use of collaborative technology to enable students from different nations to contribute to projects using virtual teams. Computer-based training (CBT) has been widely used in the business world, but is just becoming more prevalent in academia. CBT is flexible because it allows students to learn concepts outside of the classroom and the student can review a lesson as many times as necessary to learn the material.

A more formal and extensive use of new media to change learning is embodied in the concept of distance education, which is being investigated by many universities as an alternative instructional paradigm. Duke University's GEMBA program allows students to receive an MBA from Duke, regardless of where they live around the globe.[2] The material—through the use of discussion groups, e-mail, and streamed-video lectures—is delivered to the student's computer. Not only are new media changing classroom delivery, but they are also impacting access to information outside of the classroom. Online textbooks and associated testing instruments, such as *The Internet Online: A Changing World* (Fuller et al. 1998), are now available through web-based interfaces. The University of Delaware recently decided to eliminate thousands of dollars of journal subscriptions in favor of a "virtual library" (Guernsey 1997), where desired articles or books are delivered via an overnight service to the student's or the instructor's door.

It is important to note that while new computer-based technologies may be driving much of our interest in a validated reliable construct of "media use," conventional media (verbal lectures, chalkboards, overheads, etc.) cannot be excluded from our definition of "media" or our analysis. In studying the effective use of media, our interest lies not just on whether the specific media being used is computer-based, but also on "is whatever media being used inside the classroom environment being used effectively?"

The extension of teaching dimensions to include measures of active learning and media use is necessitated by what is already occurring in the classroom. Research on these dimensions is expanding beyond traditional outlets (in periodicals focused on education) to mainstream business journals. Examples of this include the *Academy of Management Journal's* (December 1997) special issue devoted to "Teaching Effectiveness in the Organizational Sciences," as well as the *MIS Quarterly's* (September 1995) special issue devoted to "Curricula and Pedagogy." In order to understand these new techniques and technologies appropriately, we must first have an accurate means of assessing their use and effectiveness in the classroom. This research is a step in that direction.

3. THEORETICAL MODEL

Based on our review of the literature (Table 1), we hypothesize that teaching activities can be broken down into seven constructs—class organization (CO), grading fairness (GF), instructor's relationship with students (RS), workload (WL), and instructor's knowledge of the material (KM), media use (MU), and active learning (AL)—and that these constructs significantly predict student perceived performance (PF).

---

[2]A description of the program, and the technology used to deliver it, can be found on the Internet at http://www.fuqua.duke.edu/programs/gemba/index.htm.

Table 1. Construct Descriptions

Construct	Description	Supporting Research
Media Use (MU)	Effectiveness in media use to communicate course concepts	• Leidner and Jarvenpaa (1993) • Alavi (1994)
Active Learning (AL)	Degree that student is involved with classroom activities; degree that the student participates versus listens to class material	• Bonwell and Eison (1991) • Meyers and Jones (1993)
Relationship with Students (RS)	Conveying a helpful attitude; treatment of students respectfully; supportive when asked questions	• Wotruba and Wright (1975) • Anonymous (1977) • Centra (1979) • Das, Frost, and Barnowe (1979) • Dickinson and Zellinger (1980)
Knowledge of the Material (KM)	Competent in field of study; knowledgeable of instructional material; ease with which the material is explained	• Wotruba and Wright (1975) • Dickinson and Zellinger (1980) • Anonymous (1977)
Course Organization (CO)	Clarity of scheduling; course design; clarity of course objectives; overall course organization	• Hildebrand, Wilson, and Dienst (1971) • Wotruba and Wright (1975) • Anonymous (1977) • Centra (1979) • Das, Frost, and Barnowe (1979)
Grading Fairness (GF)	Clarity of exams; objective grading criteria; assignments returned with a reasonable period of time	• Wotruba and Wright (1975) • Anonymous (1977) • Centra (1979) • Das, Frost, and Barnowe (1979) • Dickinson and Zellinger (1980)
Workload (WL)	Scope of material covered; appropriate time given to complete assignments; appropriate difficulty of assignments given nature of course	• Anonymous (1977) • Centra (1979)

The first four constructs (CO, GF, RS, and WL) were identified as primary categories of ratings by Centra (1973). Instructor's knowledge of material (KM) was extracted from research by Wortruba and Wright (1975) on the qualities of effective teaching. The inclusion of two additional constructs (MU and AL) follows from more recent research on the effect of new technologies (or media) that can facilitate the educational process (Leidner and Jarvenpaa 1993), and research on active learning techniques designed to increase students' involvement in the learning process (Bonwell and Eison 1991; Meyers and Jones 1993). For the dependent construct (student perceived performance), four items were created based upon student satisfaction with the instructor and course, as well as student perceived learning. The questions included in the final survey instrument are listed in Table 2.

To assess the model's level of reliability and validity, the measurement model for the exogenous constructs will first be examined. This model for the exogenous constructs is illustrated in Figure 1. The diagram uses LISREL notation to describe the constructs and their interrelationships. Lambdas (λ_{ij}, where i = question number for that construct and j = construct number) denote the factor loadings for a specific construct, and phi's indicate the covariances between the constructs (ϕ_{kl}, where k = the first construct and l = the second construct).

Table 2. Survey Constructs and Indicators

The following questions are broken down by construct. Reverse-scored questions are in **bold**.

Construct #1: Class Organization (CO)
X_{11} The course schedule changed so much that I was never sure what we were doing in class.
X_{21} The course was well designed.
X_{21} The course objectives were clearly defined.
X_{41} The course was disorganized.

Construct #2: Media Use (MU)
X_{12} The instructor used media effectively.
X_{22} The media used in this course helped me learn.
X_{32} The media used in this course helped make the course interesting.
X_{42} Media were used in this course to effectively communicate course concepts.

Construct #3: Active Learning (AL)
X_{13} The instructor promoted discussion in class.
X_{23} The instructor raised challenging questions for discussion in class.
X_{33} Instead of just listening to lectures, I was actively engaged in the learning process.
X_{43} I was more of a participant in class than an observer.

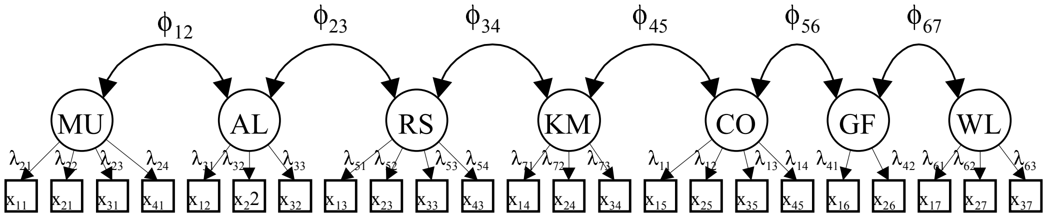
Construct #4: Grading Fairness (GF)
X_{14} The exam questions were clear and unambiguous.
X_{24} The grading in this course was fair.
X_{34} The instructor returned graded assignments within a reasonable time period.

Construct #5: Relationship with Students (RS)
X_{15} The instructor was difficult to get along with.
X_{25} The instructor provided help when asked.
X_{35} The instructor treated students with respect.
X_{45} The instructor was helpful and supportive.

Construct #6: Workload (WL)
X_{16} This course covered too much material.
X_{26} I had adequate time to complete course work.
X_{36} Assignments were unreasonably difficult.

Construct #7: Knowledge of the Material (KM)
X_{17} The instructor is an expert in his/her field.
X_{27} The instructor was very knowledgeable.
X_{37} The instructor struggled with the course material.

Construct #8: Student Perceived Performance (PF)
Y_{18} I was satisfied with this course.
Y_{28} I learned a lot from this course.
Y_{38} I didn't learn much from this course.



Note that the covariances between constructs are fully connected. Some connections are omitted for schematic clarity.

Figure 1. Measurement Model

4. RESEARCH METHODS

The hypothesized model will be examined using structural equation modeling (SEM) and ordinary least squares (OLS) regression. The measurement model will be examined using LISREL 8, a data analysis tool that utilizes maximum likelihood techniques to assess overall model fit when compared to actual data. LISREL is similar in concept to confirmatory factor analysis. An advantage of structural equation modeling is that it allows the calculation of reliability and validity metrics. Reliability levels will be calculated using the analysis of covariance structures, as described in Phillips and Bagozzi (1986) and described more completely later in this paper. Convergent validity will be determined by examining the significance of the factor loadings for each proposed construct (Bagozzi 1980). Discriminant validity tests whether two constructs are different (i.e., discriminant). The Joreskog (1974) procedure will be used to test such a hypothesis.

In early articles utilizing LISREL, the chi-square statistic was initially used to assess overall model fit (Fornell 1983; Lawley and Maxwell 1971). More recent articles (Bagozzi and Heatherton 1994; Bentler 1990; Browne and Cudeck 1989; Joreskog and Sorbom 198; Marsh 1988; McDonald and Marsh 1990), however, have suggested examining other indicators of model fit that are not as sensitive to sample size and departures from multivariate normality. The following statistics will be used, therefore, to assess the fit of the measurement model (Torkzadeh and Koufteros 1996): the chi-square to degrees of freedom ratio, the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), and the standardized root mean squared residual (RMSR).

4.1 Background on the Sample

All participants in the study were students attending classes during the first summer session of a private Southwestern university. The faculty teaching during the summer were asked to voluntarily participate in a research study investigating a new teaching evaluation instrument. Only classes within the business school were surveyed. Five business school departments participated in the research and the largest number of students took courses offered by the management department (31.9%), with accounting (28.4%) and information systems (24.2%) close behind.

Descriptive information regarding the sample is listed in Table 3. More males (53.3%) than females (43.8%) comprise the sample. As is typical of summer courses, most students took courses that are required in their field (43.3%). A smaller number took courses required for graduation, but not in their field of study (25.0%). A large percentage of the students were seniors (60.3%); the next highest percentage were juniors (18.3%). Note that while the surveys were administered largely to undergraduate students, a significant number of graduate students also participated in the research (13.6%).

4.2 Data Collection Procedures

While student evaluations are not typically administered during the summer semesters at this university, instrument administration and data collection procedures mimicked the procedures typically followed during the fall and spring semester in order to minimize data collection difficulties. Each faculty member usually reserves 15 minutes at the end of a class period toward the end of the semester and asks for a student volunteer to administer the surveys. To keep the administration of the surveys as uniform as possible, faculty were asked to read the following statement before the surveys were distributed to students:

Table 3. Descriptive Information

Gender Distribution:		
Male:	388	53.3%
Female:	319	43.8
Missing:	21	2.9
Course Type:		
Required in Field	315	43.3%
Required, out of Field	182	25.0
Elective in Field	113	15.5
Elective out of Field	91	12.5
Missing	27	3.7
Class Status		
Senior	439	60.3%
Junior	133	18.3
Graduate	99	13.6
Sophomore	30	4.1
Freshman	5	0.7
Missing	22	3.0
Class Departments		
Management	232	31.9%
Accounting	207	28.4
Information Systems	176	24.2
Finance	91	12.5
Economics	22	3.0

In an effort to improve the procedures in which faculty are evaluated, I ask that you complete this form evaluating the quality of the instruction in this course. It is important that you complete the form honestly and completely. While participation in the process is voluntary, your sincere cooperation will enable us to do a better job of evaluating and improving instruction here at [name of school omitted].

The faculty were instructed to leave immediately after reading the statement to insure that the procedure remained confidential and to minimize student apprehension.

Given the sensitive nature of performance data, both student anonymity and faculty confidentiality were stressed in this research. First, the traditional teacher evaluation instrument was designed to allow the student to anonymously rate instructors. Student anonymity was employed to increase the likelihood of honesty in student responses and decrease the likelihood of instructor retribution for poor ratings. Second, the collection of performance information in any organization is sensitive. The researchers, therefore, instituted a coding scheme to reduce their ability to trace survey results back to individual faculty members. The use of the coding scheme minimized the chance that any one person had access to both the data and the faculty names.

5. MODEL RESULTS

The measurement model was assessed using LISREL 8, a tool that utilizes maximum likelihood techniques to test the goodness of fit for a theoretical model versus actual data. Table 4 lists the results of this model under the "Model 1" column. With the exception of AGFI, the model meets or surpasses all criteria for model fit outlined in section 4 (AGFI was 0.89, while the standard is 0.90). All factor loadings are significant with $p < 0.01$; however, two indicators are below the 0.50 desired level as outlined in Phillips and Bagozzi (1986).

Table 4. Model Results

		Model 1[*]	Model 2		
Model Standards[=]: ratio < 5.0 GFI ³ 0.90 AGFI ³ 0.90 Std. RMR £ 0.05		$\chi^2_{(254)}$=836.62 ratio = 3.29 GFI = 0.91 AGFI = 0.89 Std. RMR = 0.045	$\chi^2_{(209)}$=665.20 ratio = 3.18 GFI = 0.92 GFI = 0.90 Std. RMR = 0.042		
	Question	λ_{ij}	λ_{ij}	ϵ_i	Reliability (ρ)
Class Organization	X_{11}	0.58**	0.58**	0.66	0.81
	X_{21}	0.70**	0.70**	0.51	
	X_{31}	0.81**	0.81**	0.34	
	X_{41}	0.77**	0.77**	0.41	
Media Use	X_{12}	0.78**	0.78**	0.39	0.88
	X_{22}	0.82**	0.82**	0.33	
	X_{32}	0.82**	0.82**	0.33	
	X_{42}	0.80**	0.80**	0.36	
Active Learning	X_{13}	0.76**	0.74**	0.45	0.76
	X_{23}	0.80**	0.83**	0.31	
	X_{33}	0.56**	0.57**	0.68	
	X_{43}	0.46**			
Grading Fairness	X_{14}	0.58**	0.58**	0.66	0.54
	X_{24}	0.63**	0.63**	0.60	
	X_{34}	0.44**			
Relationship with Students	X_{15}	0.52**	0.52**	0.73	0.82
	X_{25}	0.76**	0.75**	0.44	
	X_{35}	0.80**	0.80**	0.36	
	X_{45}	0.82**	0.82**	0.33	
Workload	X_{16}	0.59**	0.59**	0.65	0.64
	X_{26}	0.60**	0.60**	0.64	
	X_{36}	0.64**	0.63**	0.60	
Knowledge of Material	X_{17}	0.74**	0.74**	0.45	0.76
	X_{27}	0.86**	0.86**	0.26	
	X_{37}	0.54**	0.54**	0.71	
Performance	Y_{11}	0.84**	0.84**	0.29	0.89
	Y_{21}	0.88**	0.88**	0.23	
	Y_{31}	0.83**	0.83**	0.31	

[*] Error levels and construct reliability statistics for Model 1 were omitted for clarity. Results were comparable to Model 2 levels.
[=] See paper for the sources for the model standards.
**p < 0.01

For this reason, indicators x_{34} and x_{43} were dropped and the subsequent model was tested. The resulting model (Model 2 in Table 4) resulted in an improvement for all model criteria metrics. The model's $\chi^2_{(209)}$=665.20 results in a ratio of 3.18, far below the 5.0 recommended level. The GFI and AGFI metrics meet or exceed the 0.90 level, and the 0.042 standardized root mean squared residual indicates the model does an acceptable job of explaining the collected data. All factor loadings are significant with $p < 0.01$.

5.1 Reliability Calculations

The reliability of the constructs involves assessing the degree to which the data are measured without error and, therefore, yield consistent results. While Cronbach's alpha is frequently cited as a gauge of internal consistency, the metric requires the assumption that all indicators are equally important in explaining the construct. This assumption can easily be violated when some questions explain significantly more of the construct's variance than other questions—a frequent occurrence in behavioral research.

For this reason, internal consistency was assessed using the Analysis of Covariance Structures methodology, as described in Phillips and Bagozzi. The following equation is used to calculate the reliability of a single item (question):

$$\rho_i = \frac{\lambda_i^2}{\left(\lambda_i^2 + \epsilon_i\right)}$$

where ρ_i is the reliability of item i, λ_i is the factor loading for the i^{th} item, and ϵ_i is the error in measurement. Since the total variance for any indicator is standardized to 1.00, the reliability of any item reduces to the square of its factor loading (or lambda) and ϵ_i is $(1-\lambda_i^2)$.

The composite reliability of a construct can be determined from the following equation:

$$\rho_c = \frac{\left(\sum_{i=1}^{n} \lambda_{ic}\right)^2}{\left(\left(\sum_{i=1}^{n} \lambda_{ic}\right)^2 + \sum_{i=1}^{n} \epsilon_{ic}\right)}$$

where ρ_c is the composite reliability of construct c and n is the number of questions for construct c. The item and composite reliabilities are listed in Table 4. In all cases, the composite reliabilities are greater than 0.50, indicating that the constructs are measured reliably.

5.2 Validity

Convergent validity is the extent of agreement between two different measures of a theoretical construct. Bagozzi (1980) recommends that for convergent validity to occur, all factor loadings (i.e., lambdas) must be large and statistically significant. An examination of the factor loadings in the theoretical model demonstrates that all loadings are sufficiently large (the lowest value is 0.52). To determine whether or not the loadings are significant, Stevens (1986) recommends considering sample size in determining the significance level of the factor loading. For a sample size of approximately 700, Stevens recommends retaining a factor loading of no less than 0.20. Stevens also recommends, however, that a loading should explain a minimum of 15% variance, indicating a minimum loading of approximately 0.40. In either case, the loadings in the hypothesized model meet the recommended criteria.

Discriminant validity tests the null hypothesis that two constructs measure the same theoretical concept. Conceptually, two constructs that are different (i.e., discriminant) should have a correlation significantly different than one. Joreskog (1974) outlines a procedure to test such a hypothesis. The test requires comparing the theoretical model with a model in which the correlation between the tested constructs is constrained to one. If discriminant validity exists, the resulting model should have a χ^2 value significantly higher than the theoretical model. Since the constrained model will gain a degree of freedom with the fixed correlation, the $\chi^2_{(df=1)}$ value is tested for significance. All constructs in the model must be tested pairwise in the determination of discriminant validity. Models where the χ^2 increases more than 6.64 (the critical value for χ^2 at $\alpha = 0.01$ with $df = 1$) for the constrained constructs satisfies a test for discriminant validity for those constructs.

The values in Table 5 reflect the difference in χ^2 from the value in the theoretical model (i.e., $\chi^2_{(209)} = 665.20$) and all possible constrained models (e.g., $\chi^2_{(210)} = 675.59$ for the model where class organization and grading fairness have been constrained to one, resulting in a difference of $\chi^2_{(1)} = 10.39$). In all of the above cases, the constrained model's χ^2 is significantly higher than that of the theoretical model, indicating Joreskog's discriminant validity criterion is satisfied.

To determine the effect of the exogenous constructs on student perceived performance (PF), ordinary least squares analysis was used.[3] The stepwise procedure was followed to ensure indicators were sufficiently significant predictors of performance. The stepwise procedure will also remove an indicator if its predictive significance drops below a defined level (in this case, the 95% significance level).

Table 5. Test for Discriminant Validity:
Hypothesized Model Versus a Constrained Model

	Class Org.	Grading Fairness	Workload	Media Use	Active Learning	Student Rel.
Grading Fairness	10.39					
Workload	154.47	29.78				
Media Use	515.06	48.26	185.89			
Active Learning	268.62	60.06	238.35	423.17		
Student Relationships	365.56	28.25	196.08	744.53	343.33	
Knowledge of Material	192.42	27.89	223.00	365.02	248.61	268.55

Note: the above values reflect the difference in c^2 between the hypothesized model and a model where the correlation has been fixed to one for the indicated constructs.

Table 6. Stepwise Regression Results
Constructs Listed in Decreasing Order of Effect

Predictor	Coefficient	Standard Dev.	T-value	P-value	VIF
Constant	0.02	0.02	0.84	0.404	
CO	0.26	0.03	8.23	0.000	2.0
GF	0.22	0.03	7.50	0.000	1.7
SR	0.18	0.03	6.22	0.000	1.6
MU	0.16	0.03	6.03	0.000	1.4
AL	0.12	0.03	4.45	0.001	1.5
WL	0.08	0.03	3.26	0.001	1.3
KM	0.08	0.03	2.77	0.006	1.8

Model Fit Statistics:
s = 0.5620 R^2=67.3% R^2(adjusted) = 67% $F_{7,618} = 181.92$

[3]A causal LISREL model was completed, but the slopes of the seven independent constructs were skewed because of multicollinearity (e.g., high intercorrelation). The slopes of the exogenous variables, therefore, were calculated using OLS (ordinary least squares). Constructs were calculated by standardizing the variables and taking the arithmetic mean.

The stepwise procedure entered all variables into the regression in the order listed in Table 6; no variables were removed. The initial F-test for linearity indicates that the model is, indeed, linear ($F_{7,618} = 181.92$, $p < 0.001$). The standard error of the model is s = 0.5620, and the $R^2 = 67.3\%$, indicating the model explains approximately two-thirds of the overall variance. Note that because the data have been standardized, the coefficients in Table 6 are scaled in terms of standard deviation. The standardization also facilitates comparisons among the predictors. All hypothesized predictors of performance are highly significant at the $p < 0.01$ level. Correlation between the standardized residuals and the normal probability values was 0.98, indicating that the assumption of normal residuals is met.

6. DISCUSSION AND FUTURE RESEARCH

This study examined the dimensions of teaching within a business school environment and the strength of those dimensions in the prediction of student perceived performance. It was hypothesized that the constructs defined a priori would be significant predictors of student perceived performance. Data analysis supported the predictive ability of all the hypothesized constructs.

The strongest predictor of student perceived performance appears to be class organization. The creation of a well-defined class schedule, class objectives, and assignments may assist students in structuring their time as well as their thinking. Student may not appreciate the uncertainty of ad hoc classes; in fact, students frequently chide faculty for ambiguous assignments and not communicating clearly what is expected of them within the course. In some instances, the creation of ambiguous situations may actually enhance learning, and thus presumably instructor performance (despite the fact the students may not immediately appreciate the value of such a learning exercise). By creating a situation where the student must not only find the answer but the question, faculty can encourage unstructured thinking within the classroom.

Similar to course organization, students also appreciate the structure of fair grading practices within a course. The indicators for this construct capture students' perceptions that exam questions were clear and fair; that the course grading was objective; and that the instructor was punctual in providing students feedback on the quality of their work. All of these characteristics appear to be important in increasing students' perceptions of instructor performance.

Relationship with students was also found to be a major predictor of student perceived performance. A frequent complaint among faculty is the belief that students prefer to be entertained rather than taught. In such a situation, it seems likely that students would like their professor and, therefore, it would be easy to dismiss the strength of a student relationship construct. The questions making up this construct, however, tell a different story. This construct emphasizes the importance of helping students and treating students with respect. Previous research (Wotruba and Wright 1975) has found that instructors who create a comfortable environment also create an environment that is more conducive to learning.

The relative strength of the new constructs was encouraging. Media use was found to be fourth in importance, indicating the selection and effective use of communication tools are not only important, but may improve the students' educational experience. Leidner and Jarvenpaa (1993) found that the effective use of computer-based technology can enhance learning in the classroom. They also found, however, that how the technology was used was also important. Classes that used computers simply as a presentation device saw no advantage over overhead projectors or other traditional display media. When the technology is used as an analysis and discussion tool, however, the classes experienced more stimulating discussions that involved more complex thought processes—such as analysis—than simple processes—such as memorization. It is possible, therefore, that classes integrating new technologies (such as Internet-based discussion groups, chat sessions, and electronic brainstorming) can result in stronger teaching perceptions by students.

Active learning, another new construct focused on involving students in the classroom learning experience, was also a significant predictor of student perceived performance. The strength of the active learning construct supports a constructivist model of learning, where the classroom environment is learner-centered (O'Loughlin 1992) as opposed to an objectivist model, where the classroom is instructor-centered (Jonassen, Peck and Wilson 1999). Recent articles highlight the importance of involving students in the learning experience and demonstrate that the passive lecture environment is a poor method of communicating information. While potentially having a large impact, classes that cognitively engage students can be more difficult to conceive, plan, and implement. Students may realize this and appreciate the added attention and effort that they receive during such classes.

While significant, workload and knowledge of the material were less important predictors of student perceived performance. The positive correlation for workload may need some clarification. Classes that require more work may also enable students to increase their perceived learning. Indeed, while it is clear some students avoid classes that require much too work, it is also clear that students will not tolerate classes that require too little of them. Consistent with previous research findings, the instructor's knowledge of the material is a significant—but weak—predictor of student perceived performance. While in extreme cases students may be able to detect low domain knowledge, for the most part students are not qualified to determine whether or not an instructor is knowledgeable. This construct, therefore, probably taps into the instructor's clarity of presentation and ability to enunciate and explain important class terms and concepts.

The implications of this research are multifold. While confirming the major dimensions of teaching established in past research, the current study extends the list of teaching dimensions to include two new important factors: media use and active learning. With the accessibility and employment of a variety of new media types into the educational sector, incorporation of indicators that measure effective media use becomes crucial for a full understanding of the educational process. While the continuing development of new computer technology may be largely responsible for driving our interest in the concept of effective media use, the concept of engaging the student in the learning process (called active learning in modern literature) is as old as learning itself. Despite its obviousness, the concept of active learning is largely not included in most instruments that assess the dimensions of instruction. The inclusion of such a dimension is vital to business schools for several reasons.

First, it is believed that active learning can lead to more comprehension and recall of learning objectives. To determine when active learning techniques are desirable, however, reliable and valid methods must be developed. As an example, in many programs (particularly MBA and executive education), the use of case studies and in-class discussions are a common delivery method of educational material. Business schools need the ability to compare these types of active learning environments to more traditional lecture or non-interactive forms in order to provide adequate feedback to instructors.

Second, new computer technology is being employed with the belief that it can substitute for, or in some instances improve, the amount of learning that takes place in the classroom. One of the primary benefits posited by supporters of technology, such as multimedia computer-based-training modules (CBT), is that these customized and student-paced environments provide a more active learning environment. Future research in this area can address whether these new learning tools are doing what we think they are doing (e.g. increasing active learning), as well as their bottom line impact on student understanding and retention.

While this research helps establish active learning and media use as important new teaching dimensions, the researchers want to add a word of caution in using this instrument to evaluate faculty. This instrument was designed to assess the dimensions of activity that occur in the learning environment. It was not designed to assess the *overall performance* of the instructor or content deliverer. In other words, while it may be generally believed that increasing the amount of active learning techniques in the classroom will lead to increased student learning, it is not possible to automatically assume causality for an individual case. A particular instructor may be quite good at stimulating learning using a traditional lecture delivery. In this instance, while an active learning score may be low, student learning may be high, indicating that the instructor is being effective in the job.

In contrast, what this instrument *can* be used for is as a tool for assessing "what" an instructor is doing in a class, rather than effectiveness at doing it. As an example, if a departmental chair receives repeated student complaints about a particular instructor, an evaluation (like that developed in this research), which assesses what has occurred in the class, may give some indications as to why these reports on poor performance have surfaced. Personal interviews between the chair and faculty member can then be employed to help ferret out the true cause, or at least shed additional light on the problem. Future research that explores the relationship between these dimensions and performance is an area where much research can be done.

While the proposed model does a good job of explaining the collected data, a number of limitations should be considered. First, the data for this research were collected during the summer session of the school year. This time of year had a number of advantages, most notably that teaching evaluations are typically not conducted during the summer, thus allowing the researchers to collect data without interfering with the normal administration of university student evaluations of instruction. While providing an opportunity for data collection, student attitudes and experiences may differ in the summer when compared to a typical fall or spring semester. For example, given the rather hectic pace of summer classes, the construct *workload* may take on special significance compared to its salience during the fall or spring. In addition, it was noted that the subject pool in this study consisted of an inordinately large number of seniors and juniors completing questionnaires, and that underclassmen and graduate students comprise a small percentage of the sample. This may also raise some generalizability issues. The relative importance of a construct such as *relationship with students*, for instance, may be more important to freshmen or sophomores than juniors or seniors.

The study also explored the constructs from a "top-down" approach. The study was confirmatory in that it relied on previous research to supply important foundational constructs in the assessment of collegiate faculty. These foundational constructs act as a benchmark against which the new constructs of media use and active learning can be compared. To effectively measure active learning and media use, the literature was examined to determine the appropriate questions to best measure these new constructs. While this approach was efficient in allowing the researchers to study the constructs of interest within the constrained time frame, more research is necessary to investigate the constructs to determine their underlying components.

6.1 Conclusion

Given lack of attention by business schools to their performance evaluation systems, the findings of this research can be interpreted two ways. First, that research is more highly valued than teaching. Second, that we simply have no valid and reliable measure of teaching, thus universities hire using the only measure they can, i.e., research ability. If the latter is true, we need better measures both to reward those faculty doing an excellent job of teaching and to assist those faculty who need help. While corporations are striving to develop more sophisticated techniques in performance evaluation, such as 360-degree feedback and multirater assessment (O'Reilly and Furth 1994), universities have not kept pace. However, given new market pressures on these same universities, it is likely that the winners in the 21st century will have addressed this need.

7. REFERENCES

Alavi, M. "Computer-Mediated Collaborative Learning: An Empirical Evaluation," *MIS Quarterly*, 1994, pp.159-174.

Anonymous. *ICES Item Catalog*, Newsletter No. 1, Champaign-Urbana, IL: Office of Instructional Resources, University of Illinois, 1997.

Astin, A. W. *Achieving Educational Excellence*, San Francisco: Jossey-Bass, 1985.

Bagozzi, R. *Causal Modeling in Marketing*, New York: Wiley & Sons, 1980.

Bagozzi, R. andHeatherton, T. "A= General Approach to Representing Multifaceted Personality Constructs: Application to State Self-esteem," *Structural Equation Modeling* (1:1), 1994, pp. pp. 35-67.

Bentler, P. "Comparative Fit Indices in Structural Models," *Psychological Bulletin* (107), 1990, pp. 238-246.

Bonwell, C. C., and Eison, J. *Report No. ASHE-ERIC Higher Education Report 1*, Washington, DC: The George Washington University School of Education and Human Development, 1991.

Browne, M., and Cudeck, R. "Single-sample Cross-validation Indices for Covariance Structures," *Multivariate Behavioral Research* (24), 1989, pp. 445-455.

Centra, J. A. "Effectiveness of Student Feedback in Modifying College Instruction," *Journal of Educational Psychology* (65:3), 1973, pp. 395-401.

Chickering, A. W., and Gamson, Z. F. "Seven Principles for Good Practice," *AAHE Bulletin*, 39, 3-7.

Coffman, W. "Determining Students' Concepts of Effective Teaching from Their Ratings of Instructors," *Journal of Educational Psychology* (45), 1954, pp. 277-285.

Cross, P. K. "Teaching for Learning," *AAHE Bulletin* (39:283), 1987, pp. 3-7.

Das, H.; Frost, P.; and Barnowe, J. "Behaviorally Anchored Scales for Assessing Behavioral Science Teaching," *Canadian Journal of Behavioral Science* (11), 1979, pp. 79-88.

Densford, L. E. "Call it the Airline with an Attitude," *Corporate University Review: The Magazine About Organizational Learning and Performance* (5:5), 1997 (available: http://www.traininguniversity.com/magazine/current/ [December 12, 1997]).

Dickinson, T., and Zellinger, P. "A Comparison of the Behaviorally Anchored Rating and Mixed Standard Scale Formats," *Journal of Applied Psychology* (65), 1980, pp. 147-154.

Erickson, S. C. *The Essence of Good Teaching: Helping Students Learn and Remember What They Learn*, San Francisco: Jossey-Bass, 1984.

Fornell, C. "Issues in the Application of Covariance Structure Analysis: A Comment," *Journal of Consumer Research* (9), 1983, pp. 443-448.

Fuller, M. A.; Serva, M. A.; Trower, J.; and Willis, G. W. *The Internet Online: A Changing World*, New York: McGraw-Hill, 1998 (available online at http://www.mhhe.com/cit/net/trower1 [May 1998]).
Grussing, P. G. "Sources of Error in Student Evaluation of Teaching," *American Journal of Pharmaceutical Education* (58), Fall 1994, pp. 316-318.
Grussing, P. G.; Valuck, R. J.; and Williams, R. G. "Development and Validation of Behaviorally-anchored Rating Scales for Student Evaluation of Pharmacy Instruction," *American Journal of Pharmaceutical Education* (58), Spring 1994, pp. 25-37.
Guernsey, L. "University of Delaware Library Uses Web to Retrieve Journal Articles," *The Chronicle of Higher Education*, 1997, p. A24.
Hildebrand, M.; Wilson, R.; and Dienst, E. *Evaluating University Teaching*, Berkeley: University of California Center for Research and Development in Higher Education, 1971.
Hodgson, T. F. *The General and Primary Factors in Student Evaluation of Teaching Ability*, Seattle: University of Washington Press, 1958.
Ives, B., and Jarvenpaa, S. L. "Will the Internet Revolutionize Business Education and Research?" *Sloan Management Review* (37:3), 1996, p. 33.
Jonassen, D. H.; Peck, K.; and Wilson, B. G. *Learning with Technology: A Constructivist Perspective*, Upper Saddle River, NJ: Merrill,
Joreskog, K. G. "Analyzing Psychological Data by Structural Analysis of Covariance Matrices," in *Contemporary Developments in Mathematical Psychology: Measurement, Psychophysics, and Neural Information Processing, Volume 2,* R. C. Atkinson, D. H. Krantz, R. D. Luce, and P. Suppes (eds.), San Francisco: Freeman, 1974, pp. 1-56.
Joreskog, K., and Sorbom, D. *LISREL VII User's Guide*, Mooresville, IN: Scientific Software, Inc., 1989.
Lawley, D., and Maxwell, A. *Factor Analysis as a Statistical Method*, London: Butterworth, 1971.
Leidner, D. E., and Jarvenpaa, S. L. "The Information Age Confronts Education: Case Studies on Electronic Classrooms," *Information Systems Research* (41:1), 1993, pp. 24-54.
Marsh. H. W. "Goodness-of-fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size," *Psychological Bulletin* (103:3), 1988, pp. 391-410.
McDonald, R., and Marsh, H. "Choosing a Multivariate Model: Noncentrality and Goodness-of-fit," *Psychological Bulletin* (107), 1990, pp. 247-255.
Meyers, C., and Jones, T. B. *Promoting Active Learning: Strategies for the Classroom*, San Francisco: Jossey-Bass Publishers, 1993.
O'Loughlin, M. "Rethinking Science Education: Beyond Piagetian Constructivism toward a Sociocultural Model of Teaching and Learning," *Journal of Research in Science Teaching* (29:8), 1992, pp. 791-820.
O'Reilly, B, and Furth, J. "Managing 360 Feedback can Change Your Life: What Your Boss, Your Peers, and Your Subordinates Really Think of You May Sting, but Facing the Truth Can Also Make You a Better Manager," *Fortune* (10:17), 1994, p. 93.
Phillips, L., AND Bagozzi, R. "On Measuring Organizational Properties of Distribution Channels: Methodological Issues in the Use of Key Informants," *Research in Marketing* (8), 1986, pp. 313-369.
Robbins, T. "Meaningfulness and Ccommunity in the Classroom: The Role of Teamwork in Business Education," *Journal of Education for Business* (69), 1994, pp. 312-317.
Seldin, P. *Changing Practices in Faculty Evaluation*, San Francisco: Jossey-Bass, 1984.
Slavin, R. E. *Cooperative Learning: Theory, Research, and Practice*, Englewood Cliffs, NJ: Prentice Hall, 1990.
Stevens, J. *Applied Multivariate Statistics for the Social Sciences*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.
Tagomori, H. T., and Bishop, L. A. "Student Evaluation of Teaching: Flaws in the Instruments," *The NEA Higher Education Journal* (11:1), 1995, pp. 63-78.
Torkzadeh, G., and Koufteros, X. "Structural Equation Modeling for Confirming the Self-efficacy Instrument," *Proceedings of the 1996 Decision Sciences Institute*, 1996.
Van Ort, S. "Developing a System for Documenting Teaching Effectiveness," *Journal of Nursing Education* (22:8), 1983, pp. 324-328.
Wotruba, T. R., and Wright, P. L. "How to Develop a Teacher-Rating Instrument," *Journal of Higher Education* (XLVI:6), 1975, pp. 653-663.