

STUDENT PERCEPTIONS OF FACULTY INSTRUCTIONAL VALUE-ADDED: A NEW MEASURE AND EXPLORATORY EMPIRICAL EVIDENCE

Timothy J. Fogarty

*Weatherhead School of Management
Case Western Reserve University
Cleveland, Ohio
USA*

Brian Hogan

*College of Business Administration
Northeastern University
Boston, Massachusetts
USA*

ABSTRACT

Students react to two basic things when they are asked to rate a college course. Their ratings will reflect a certain response to the course content and to the method in which that content was delivered by a faculty person. We should expect that the resulting opinion of the teacher exists somewhat independent of the value that students perceive in the content of the courses that are taught. This paper defines this difference as a teacher's instructional value-added. That some teachers are more successful than others in impressing students is difficult to deny. However, little is known about the nature of this increment. Using data from one school, the paper shows how instructional value-added perceived by students is distributed by discipline, by level, and by individual. Separate results are also provided for accounting classes. Suggestions for future research involving the instructional value-added construct are made as part of our continuing effort to understand and evaluate post-secondary instruction.

Key words: Student Evaluation of Teaching, Teaching Performance, Student Perceptions

Data availability: Data used in this study can be obtained by contacting the second author at b.hogan@neu.edu

INTRODUCTION

Qualitative studies of academics show that desire to be a good teacher is a noble aspiration that is both widespread and genuine (e.g., Clark, 1987). Despite considerable attention on the research activities of the modern professoriate, quality teaching remains the central component of the role, as it is understood by most of academe's internal and external constituents.

Professors are provided teaching assignments in subject matter in which they have some degree of expertise. However, most faculty do not unilaterally control the curriculum. In an ideal world, all faculty could convince all students that any subject matter in the curriculum is critical to them and an equally valuable increment to the accomplishment of their career and personal objectives. In a realistic world, all courses are not equally valued by students. Instructors assigned to difficult courses that lack an immediate and compelling connection to student lives may appear to be less than good teachers, merely because they have been unable to overcome the inherent limitation of the material. Others, fortunate to teach material that possesses elements likely to be better received, may benefit from a halo effect and therefore have their teaching overly praised. This less than level playing field complicates our ability to appreciate the efforts of faculty in the classroom.

This paper is premised on the importance of better understanding how students perceive the contributions to learning made by their faculty. For this purpose, the paper makes the case that it is necessary to control for variation attributable to the subject matter. When one does, that which remains can be called the teacher's instructional value-added. Although instructional value-added is a somewhat ambitious and multifaceted term, its tentative identification permits important avenues of inquiry into higher education. Using student ratings data from one highly ranked private business school, this paper describes dimensions of this exploratory construct. The data suggests that instructional value-added is a stable construct that can inform us about our students' worldview as well as our faculty's efforts.

This paper uses three subsequent sections. The first motivates the inquiry and reviews the related literature. The second identifies specific research questions and provides a method to test them. The final section describes the results and offers a discussion of their implications and limitations.

BACKGROUND AND LITERATURE REVIEW

The demand for accountability serves as a sign of the times. Institutions of higher education, especially those that are supported by taxpayers, have been no exception. The interest in how public and private money is being spent translates into many rather unprecedented questions that are now being asked about how colleges and universities operate. For example, external parties now demand more information about the package of services provided to students, the elements of curricula of study and how faculty spend their time. In this tradition, various groups have asserted their rights as consumers of higher education (see Buckmaster and Craig, 2000). Among other effects, these developments have increased the interest in the caliber of teaching efforts by university faculty.

At the same time, changes have occurred in the accreditation of higher education organizations. The new thinking that has altered accreditation from an assessment of critical input resources to the execution and delivery of valued outcomes has heightened the inquiry into teaching practices for many schools. Under mission-based accreditation systems, schools can select a mixture of intellectual contributions and teaching effectiveness results. Schools that are not known for the research conducted by their business faculty can put more emphasis on the impact of their

instructional efforts. At the same time, even high-powered research schools must demonstrate that they have a systematic approach to the continuous improvement of learning. In broad terms, more pressure is put on individual faculty to adhere to best practices and be more involved in efforts to contribute to institutional teaching goals (see Calderon et al. 1997; Bailey and Bentz, 1991) across higher education.

Although the extent to which a university actively rewards quality teaching will depend upon its mission elements (see Street et al. 1993), no school can be openly hostile or even indifferent to teaching. Private schools that charge premium tuition rates often feel that they must justify this situation with exceptionally good teaching. Research schools routinely profess that a faculty actively engaged in the creation of knowledge delivers qualitatively better teaching value. At the same time, schools without a strong research culture must depend even more exclusively upon their teaching prowess. Often, the logic that sustains any claim to distinction is that teachers at these schools can devote themselves more exclusively to their instructional work. Therefore, rewarding and encouraging good teaching is institutionally rational and appropriate in all sectors of the academy.

How well a faculty person's teaching is received by students constitutes an element of the trajectory of that individual's career. Even if "publish or perish" suggests that research productivity has become increasingly important to faculty, schools cannot be indifferent to relative teaching success. To the extent that teaching matters to promotion and tenure decisions, this dimension has increasingly been reduced to student evaluations metrics (Seldin, 1984; Carruth and Carruth, 1997; Raghunandan et al. 1999). Teaching abilities, as made objective in this manner, also influence academic labor market results (Lewis, 1996).

Ironically, inadequate attention has been directly focused upon the caliber of teaching by those most immediately involved. Accordingly, Reckers (1996) points to the need to acquaint faculty and administrators with the "basic production factors" that create value in the education market. This redirection may be needed because of the pervasive lack of formal training in educational methods provided to university faculty (see Stevens and Stevens, 1992). At the same time, students are perhaps blissfully ignorant about the circumstances that complicate the lives and work of those who teach at the university.

Almost every school has instituted some procedures to evaluate the quality of its teaching. The most obvious approach is to systematically collect and tabulate student perceptions. Although student abilities to understand quality teaching might be limited, their intimate proximity to the delivery of these services entitles them to a voice in instructional evaluation. Looking to students for this information is also recommended to schools by virtue of its convenience.

The circumstances outlined above suggest that teaching evaluations contribute to institutional goals, and therefore make sense from that perspective. Teaching evaluations provided by students can also be valuable for faculty members. In this context, faculty members who exceed the formal expectations of their roles can be identified as superior, and rewarded by more than those who were fortunate enough to have taken their classes. At the other end of the spectrum, student evaluations that suggest poor faculty performance can provide a faculty person with information that might lead to new approaches and altered teaching methods.

A variety of complex reasons underlies the fact that teaching evaluations are taken very seriously in most sectors of the academy. In several ways, the need for a "bottom line" which is measurable and not entirely dependent upon the context of its production explains such an institutional position. Whereas department chairpersons might be able to read the full evaluation (including narrative comments), and therefore might develop a richer interpretation of its meaning,

a numerical evaluation is capable of a life of its own outside these circumstances. Teaching evaluations that are reducible to specific positions on Likert scales can be averaged across time and across teaching assignments to get an approximation of what is generally true. These numbers also transcend institutions, becoming part of the credentials of the teacher, as he or she attempts to move from post to post. Although everyone knows that teaching is highly nuanced and variegated, the power of the objective teaching evaluation is that such complexity progressively is eliminated. Arguably, all the detail that cannot be compared is subsumed by these powerful numerical scales.

The existence of teaching evaluations does not offer a panacea. Deans continue to worry about problems in evaluating faculty (Seldin, 1984) and identify the teaching component of academic work to be more important to their organizations than research (Wisdom and Teer, 1990). In the short term, when evaluation remains imperfect, the acceptance of these instruments may encourage teachers to be risk-averse pertaining to the incorporation of paradigm shifting material (Dopuch, 1989) and be inclined toward a weakening of student competency assessment (Wallace and Wallace, 1998).

Some have sought to appreciate teaching evaluations in a broader framework. However, considerable variation exists in the nature of such a context. These include disciplinary differences (D'Onofrio et al. 1988), practitioner expectations (Meyer and Titard, 2000) and the foibles of the customer-driven market (Dopuch, 1989). A more systematic effort has attempted to re-introduce degrees of qualitative richness and idiosyncratic circumstances to a numerical system through the construction of "teaching portfolios." This attempt essentially frames the quantitative results with more of an examination of the efforts expended and the specific outcomes sought by teachers (see Green et al. 1999a).

Ironically, faculty research productivity has been the single most often studied outcome to which student-provided teaching evaluation has been put in the educational literature (see Hattie and Marsh, 1996). Apparently, the belief that teaching effectiveness and research productivity are mutually exclusive is so powerful that it persists despite considerable evidence to the contrary (e.g., Centra, 1983; Feldman, 1987). Disciplines with a cohesive external constituency of practitioners apparently are particularly prone to the misapprehension that research is a preoccupation fatal to the interests of employers in the production of a talented group of future recruits (see Benke and Roof, 1990; Bell et al. 1993). Evidence on the increasing dependency of promotion and salary to research productivity (see Read et al. 1998; Root, 1987) may have resulted in additional hours devoted to scholarship (Henke, 1998) but has not eroded the caliber of teaching efforts, at least as such is perceived by students. Thus, teaching evaluations cannot be seen as a mere appendage to the research evolution of the business school faculty.

Teaching evaluations are used by nearly all business schools (Calderon and Green, 1997). Even when offered richer opportunities to voice their opinions, students believe that these traditional assessment devices should be retained (Harwood, 1999). Students profess that the feedback these instruments provide contributes to the improvement of teaching and intelligent staffing decisions (Chen and Hoshower, 1998). Faculty also depend upon information contained in these instruments to establish the trust in, and control over, their work that is important to their satisfaction (Cares and Blackburn, 1978). Nonetheless, they also have been critiqued. For example, Wallace and Wallace (1998) suggest that, as a "happiness index," students evaluations can be manipulated by faculty and may penalize those who have rigorous academic expectations. To some extent, students that do not recognize good teaching may be asked by these evaluations for "more than they can know" (Nisbett and Wilson, 1977). Ironically, schools that place less weight upon teaching in promotion decisions

tend to rely more upon formal student evaluations (Raghunandan et al. 1999). Albeit imperfect, Likert-scaled teaching evaluations are the most cost effective and reliable instrumentation about teaching outcomes the business academy is likely to have in widespread application over the foreseeable future.

Toward a New Approach

The strong analytical approach made possible by Likert-scaled teaching evaluation may overreach that which we need to know. Just as good teaching cannot be reduced to technique, the evaluation of good teaching cannot be absolutely pinpointed to a position of a single scale. A more diffuse approach that is not so tied to course content must be devised to capture what Palmer (1998) calls the “identity and integrity” of the teacher.

At the same time, the wealth of data created by extant student evaluations of teaching is difficult to ignore. These evaluations contain clear ideas about what students value (see Osborne and Lukshin, 2000) that would be difficult to replicate. Student evaluations possess an immediacy not found in other approaches (e.g., Atkinson and Delamont, 1990). That these instruments can be administered on a sufficiently broad basis exhibits a scalability not obtainable through more customized approaches.

Although by no means comprehensive, student evaluation instrumentation represents considerable collective academic effort over many years to assess all that a teacher does that bears upon the proper instructional objective. Assessments of reliability and validity have generally been positive (e.g., Wright et al. 1984), although this conclusion does not hold true for each and every item that these instruments tend to measure (Green et al. 1999b). One should also not forget that the data produced by these evaluations may have more bearing upon the quality of teaching than on the extent or degree of learning.

For these purposes, the paper proposes the construction of a singular measure of teacher instructional value-added from the conventional metrics of student-provided data. This approach controls for the variation in student interest in subject matter and in student disagreement over the meaning to give the absolute scores (Ketler et al. 2000). As such, the new measure offers a more holistic impression of student evaluation of teaching efforts.

Student evaluations contain a wealth of detailed information about highly specific instructor behaviors (e.g., punctuality, courteousness, accessibility). They often seek reactions to very particular aspects of the course and the materials used therein (e.g., fairness of tests, adherence to syllabus, quality of the textbook). However, the importance of these specific items to the overall impression of teaching is debatable. Exactly how superior utilization of one dimension compensates for, or supplements another, is unknown. For example, how would students balance that the instructor reliably held office hours but did not always answer questions in a positive manner? To avoid these difficult aggregation issues, and to bypass items with questionable reliability and generalizability, this study focuses upon the two summative measures that transcend the clutter of inquiries into detailed technique, and specific teacher behavior. Specifically, students invariably are asked to provide a singular measure of the quality of the course and the instructor. Both of these measures should be the product of everything that the rater believed to be of importance about the course and its instructor.

Many previous studies have identified the unique importance of these two “bottom line” items (e.g., Wachtel, 1998; Langbein, 1994). To do otherwise ignores the fact that effective teaching can be done in a wide variety of ways and can use a considerably different set of materials. How

techniques and strategies of instruction combine to influence overall student judgments has been studied (e.g., Marsh and Bailey, 1993; Ainger and Thum, 1986).

The unusual aspect of this study is to insist that the difference between these two end points is itself a unique conclusion. Specifically, the difference between the teaching rating and the course rating is defined as instructional value-added. An instructor who produces a personal ranking in excess of the ranking of the material that he or she teaches adds to the overall experience by the power of superior energy or enthusiasm or personality. This can be called positive instructional value-added. In quite a literal sense, the assessment of the instructor is benchmarked against the student's reaction to the substance of the course.

When students are asked to provide a summative judgment about a course, they consider many things. These range from the caliber of the reading materials to the relevance of the subject matter to their career projections. They might select a point on the Likert scale that expresses how well-spent the time expanded on the course has been. Immediately thereafter, confronted with a similar question asking them to extract the contribution of the instructor, students will anchor around their course rating. If the instructive efforts were a positive aspect within the mix of course elements, value-added will be signified with a selection in excess of that given to the course as a whole. A course that is rated as more valuable or important than the teacher's efforts suggests that the teacher has underperformed the possibilities of the material. This difference can be termed somewhat oxymoronically, negative instructional value-added. Such a rating communicates that given important or interesting material to teach, the instructor was unable to add further utility. Again, the course sets a visible benchmark that could be seen as the potential that the course material made available to the instructor.

Value-added connotes the expectation that a layer of incremental professional servicing exists in education. In either direction, the differencing technique wherein the course rating is the touchstone allows the teaching component of the course to be roughly isolated. This technique produces supplemental data to the absolute positioning of the student evaluation of course and instructor.

RESEARCH QUESTIONS AND METHOD

Since instructional value-added is a construct about which little is known, some basic descriptive work must precede the test of any research questions. For example, although instructional value-added is expected to be normally positive, the frequency of the positive and negative directions is unknown. Whereas the evaluation of a human being will probably be more generous than that of theoretical material, the occurrence of the opposite positioning could range from very rare to quite common. Furthermore, the absolute magnitude of instructional value-added needs to be reported. Although it could be expected that course evaluations and teaching evaluations would track together, how closely this happens is an interesting question.

A statistical property of the value-added calculation points out that instructional value-added should not be equated with effectiveness. Likert scaling can be expected to induce some ceiling effects that would mute severe deviations. For example, a teacher with a course ranked 5.0 on a 5-point Likert scale who also gets a 5.0 on the overall teacher rating would appear to have no instructional value-added ($5.0 - 5.0 = 0$). This situation is the epitome of exceptional faculty work that merits considerable praise. The possibility of this situation requires one to partially disassociate the instructional value-added concept from the effectiveness of the concept. Value-added is meant

to supplement, rather than to replace, concern with the absolute position achieved on the Likert scales.

Although faculty members routinely teach different classes, their success with these assignments might not be a constant. They may greatly prefer one class over others, creating motivational differences that will show up in teaching evaluations. The academic life tends to allow people to make somewhat unilateral and distinct choices about what should be accomplished (Jackson, 1996). As such, instructional value-added may be context-specific, especially if it adheres more to some subject matter than to others. If instructional value-added is more in the nature of a personality attribute, it should manifest itself relatively invariably across different courses. A conservative estimation also would suggest that the instructional value-added of an instructor should be more variable across different classes than it is upon repeat offerings of a singular course.

Class Size

One very visible way in which teaching assignments differ is in the number of students. Large classes put pressure on faculty to resort to more formal methods of instructional delivery (e.g., lecture). They also tend to tip student assessment practices toward that which can grade via automated tools (e.g., multiple choice). Faculty members invariably complain about large classes, not only because of the extra work involved, but also because of what most perceive as constrained pedagogical choices.

Teaching is a highly charged emotional environment that could involve dynamic interaction between faculty and students. The potential for this engagement may expose a larger range of a teacher's abilities. *Ceteris paribus*, smaller classes allow for more of a faculty person's personality as an educator to be known (McKeachie, 1980). Contrariwise, larger classes may induce a more formalized instructional motif (see Light, 2001) that does not allow a unique instructional value-added effort to be evident. A strong size effect would render instructional value-added less of an indication of teaching skill than a reflection of the happenstance of scheduling.

Smaller classes invite teachers to invoke a pedagogy that is qualitatively different than can be done in larger settings. In smaller classes, instruction can be more personalized and less of a "show" that needs to be consumed on a "one size fits all basis." If a teacher takes advantage of the opportunities to customize the student experience, made possible by fewer students, the value-added proposition perceived by students might become more apparent to students. This prospect underlies the following.

RQ1: Faculty members teaching smaller classes will be perceived as having more positive instructional value-added.

Subjects possess varying degrees of inherent interest to students. This may reflect a collective belief that one set of phenomena is more valuable than another, or it may express an emotional reaction to a subject's raw materials. Subjects also vary in the demands they place on the educational backgrounds of students. For example, even within the narrow scope of a business school, students might perceive finance to be more important as a path to personal wealth than human relations, marketing to be "sexier" than accounting, and organizational behavior to be less math-intensive (and therefore more hospitable) than operations research. Although disciplines in the business school vary in their reception by students, the faculty may be less patterned by their affiliation. Each disciplines' instructors should have varying pathways to escape or outperform the

limitations of their discipline. Each area probably will have its high instructional value-added faculty and its lesser instructional value-added faculty. If the motivation to deliver a value-added classroom experience is personal in nature, not much reason exists to believe that strong disciplinary differences should exist.

Very little is known about the socialization of faculty, especially as it pertains to non-research roles. The extent to which disciplines vary in the degree they stress creative and energetic teaching is unknown. At some schools, some departments may have an ethos that disproportionately values teaching success. This could be attributable to the influence of a single persuasive role model. Thus, priorities of some departments may lead to the purposeful recruitment of instructors believed to be “good in the classroom” at the margin, and may further their distinctiveness from other academic departments.

The instructional value-added of the instructors of some disciplines may be irretrievably affected by a negative stereotype not shared by other disciplines (Mladenovic, 2000). In their evaluations of courses, some students may be unable to get beyond the expectations formed by these cultural images. Other disciplines are less known to students and therefore do not carry such baggage. The same teaching performance may therefore have much different benchmarks formed by the outlines of the disciplines. Although *ex ante* possibilities of difference are plausible, insufficient guidance exists to predict a direction.

RQ2: Instructional value-added will vary across academic disciplines.

Briscoe et al. (1998) suggest that student evaluations of both teaching and courses decline, in absolute terms, as the level of the course increases. These authors argue that students have an adverse reaction to the higher degrees of abstraction and uncertainty to which they are more exposed in higher-level courses, and therefore rate upper level courses lower. This analysis did not include lower division courses, and used rather unusual evaluative questions. Bailey et al. (2000) also find that course level differences exist in the relationship between the teaching evaluation’s individual items and the overall instructor evaluation score. Even if this association is true, there is less reason to suspect that the rating of instructors should be suppressed in upper level courses. Contrariwise, students might grow more appreciative of the instructor’s mastery as the material becomes increasing esoteric and difficult to appreciate through the text book medium. However, increased familiarity with the college experience may allow students to become more discriminating over time about what constitutes good teaching. Whether course rating declines faster than instructional ratings is an empirical question.

The comparison of student perceptions across the typical curriculum involves the growing maturation of educational appreciation. Students taking lower level classes are typically new to the university. Many are searching for a field of study and therefore unable to accurately gauge the relevance of their classes. Those taking upper division classes tend to be seasoned veterans who have made a commitment to a major. In between these points, one should expect changes in the average reaction to the subject matter contained in college courses. One can also imagine changes in appreciation for the people who teach. Since value-added is affected by the change in one measure relative to the change in the other, the balance is unpredictable. Without much previous research, precise expectations would not be prudent. Therefore:

RQ3: Instructional value-added will vary with the level of the course taught.

Schools offer multiple programs to a variety of audiences. These programs differ in their visibility and in their importance to the strategic objectives of the school. Although all programs would benefit from good teaching, the definition of instructional effectiveness might reflect the varying pedagogical ambitions of programs and the base level motivation of enrolled students. For example, undergraduate students might tend to appreciate someone who had the ability and motivation to elaborate the more difficult concepts. MBA students might prefer someone who possessed “real world” experience that better contextualized the course content. Doctoral students might tend to find more value in an instructor who possessed a publication reputation and was willing to share insider techniques of the academic trade.

In addition to appreciating the role of programmatic purposes, the various courses of study also presume a certain type of student. Varying degrees of pre-selection and pre-socialization are built into the student populations that define the likely constituents of programs. To the extent that this selectivity shapes expectations about educational experiences, it may affect measured instructional value-added. For these purposes, the undergraduate program may be the most heterogeneous in student background and motivations. Graduate students tend to be less diverse, especially as one narrows the field by discipline. Graduate students may have more precise ideas about what constitutes value-added instruction, and be more aware of the employment contingencies of these efforts. On the other hand, graduate program students’ self-selection into their fields may create a halo effect for the material delivered to students in programs that are more heavily steeped in a particular discipline. For example, the appreciation for accounting theory should be stronger in Masters of Accountancy classes than in MBA courses, even though both are at the graduate level. Thus, along the same lines as the previous research question:

RQ4: Instructional value-added will vary across educational programs.

In total, the research questions that have been presented offer some elementary information about the deployment of a scarce resource. No school ever has an excess of faculty uniquely skilled at transcending the inherent limitations of their subject matter, and adding value to the courses that they teach. To the extent that this translates into an additional increment of student motivation and satisfaction, more knowledge about the distribution of instructional value-added is very important. The research questions posed, together with other descriptive information, creates tests to offer evidence on whether instructional value-added is a personal attribute of certain individuals or is a product of how institutions allocate instructional resources.

Method

Course evaluation data over the eight year period between 1997 and 2004 were obtained from a school of management. This encompassed the efforts of seven departments distributed over a variety of programs that ranged from undergraduate degrees to doctoral degrees. A total of 2,510 sections of 209 different classes were evaluated by students. The university that houses the business school is usually ranked by the media as one of the best forty in the USA. The business programs have received similar accolades from the financial press. This school is a private research-focused entity, with the business school and the accounting department holding separate AACSB accreditation.

The university in question, as a private elite institution, prides itself on the quality of its faculty and their teaching abilities. As an economic equilibrium, the proof of these claims must

reside in the school's ability to sustain premium tuition rates. However, as a research school, teaching is not the exclusive focus, nor is teaching excellence the dominant value at this institution. Since neither the university nor the school has adopted any mandatory element (such as service learning or writing across the curriculum) the instructor's will is relatively supreme in the design and conduct of classes. Academic departments have also left faculty to their own devices in their teaching efforts.

Although only a single school is included, the vast number of courses that are embraced serves as a compensation. The large number of observations increases the confidence that there is no systematic bias coming from the way the sample is drawn. With regard to the evaluation of the teaching environment created by the instructor above and beyond that which relates to the subject matter, each course offering could be considered its own sample. Thus, the advantages of multi-sample educational data collection are achieved (Kalbers and Weinstein, 1999).

RESULTS AND DISCUSSION

Descriptive Information

The average class in the database was rated as a 4.10 (on a 5 point scale). This rating varied from a high of 5.00 in several classes across several departments to a low of 1.60 in a course in Operations Management. The measures of central tendency suggest that students tend to believe that most of their courses provide them considerable value, as judged on an absolute basis.

Students are also favorably disposed toward instructors, as judged by this data. The average rating for instructors is 4.22 (also on a 5 point scale). These ratings also demonstrated variation across the three major educational programs (undergraduate = 4.01, masters = 4.08, doctoral = 4.46). The absolute position of the teacher on the scale is consistent with the time-honored observation that students hold a generally favorable impression of faculty (Lewis, 1968). The instructor rating variable had much higher variances (0.303) than the course rating (0.272).

The above two paragraphs point toward the conclusion that instructional value-added, as defined herein, tends to be positive. In fact, the teaching rating exceeded the course rating in 71.20% of the classes in the database. On average, the deviation of the instructor ranking above the course ranking was 0.12 of a point on the Likert scale. The correlation between the two scales was 0.46, indicating that the two measures track fairly close but are not identical.

The absolute amount of instructional value-added may at first seem to be a minor matter. However, a fuller appreciation for its magnitude can be achieved by more closely examining the distribution of course evaluations. Table 1 contains the frequency array of this data. Given the range permitted by a 5 point scale, course evaluations are tightly concentrated. Nearly 40% of these rankings fall within the half point between 4.00 and 4.50, and over 54% fall within the .75 of one point between 3.75 and 4.50. Teacher evaluations are slightly less concentrated with only 35% between 4.00 and 4.50 and 47% between 3.75 and 4.50. Extremely high (>4.90) and extremely low (<2.50) scores are unusual. In such an environment, the degree of instructional value-added appears to be quite consequential in that it can break a person out of the pack with regard to the judgments that can be made about relative teaching abilities. The paucity of extremely high course evaluations, combined with the modest degree of measured instructional value-added, also limits the importance of ceiling effects.

Whether or not the spread between the rankings is a product of different course offerings can be evaluated by examining the variance of the evaluation data by instructor. The data suggest that the average variation of the instructional value-added factor across multiple courses taught by the

TABLE 1

Distribution of Course and Faculty Evaluations

<u>Range</u>	<u>Course Evaluation Frequency</u>	<u>Faculty Evaluation Frequency</u>	<u>Range</u>	<u>Course Evaluation Frequency</u>	<u>Faculty Evaluation Frequency</u>
0.00 - 2.00	2	2	4.21 - 4.30	197	195
2.00 - 2.50	16	27	4.31 - 4.40	196	210
2.51 - 3.00	81	59	4.41 - 4.50	187	227
3.01 - 3.50	219	183	4.51 - 4.60	116	130
3.51 - 3.75	264	189	4.61 - 4.70	158	204
3.76 - 4.00	392	302	4.71 - 4.80	142	217
4.01 - 4.10	176	111	4.81 - 4.90	89	161
4.11 - 4.20	210	146	4.91 - 5.00	65	147
Total-Column	1,360	1,019		1,150	1,491
Total-All Columns				2,510	2,510

same instructor is relatively low (0.17). This result is less than the average variation across multiple sections of the same course taught by different instructors (0.24). Therefore, instructional value-added appears to adhere more to the individual faculty member than to the specific activity upon which he or she might be engaged.

The distribution of instructional value-added is also an interesting descriptive element. Table 2 shows that instructional value-added ranges from a high of 1.50 to a low of - 0.83. More than a third (38%) of the observations fall within the modest instructional value-added range of 0.00 to 0.19. Another large group (26%) could be characterized as providing moderate levels of instructional value-added (falling between 0.20 and 0.39). Although 28.8% of the classes had course ratings in excess of instructor ratings, most (60.2%) of these cases were within 0.10 or less of each other. Whereas teachers generally are more appreciated than their material, extreme deviations in the evaluation of these two are exceptional. In total, the majority (62.0%) of the scores were within 0.20 of each other.

Research Question Evidence

Although this work is exploratory, several specific research questions were posed. The first ponders the extent to which instructional value-added is more likely to emerge in a smaller class. Table 3 shows that no apparent pattern exists when class size is divided into five student population intervals. Testing this relationship with OLS shows that instructional value-added is not related (at $p < .05$) to class size.

In results not shown, class size was recoded as a non-continuous variable that divided classes into small, medium, and large, using breaking points that divided the sample of classes into approximate thirds. The results did not change. Instructional value-added is not associated with class size ($p > .05$). In a third approach, the class size data was divided into two groups where 30 or fewer students defined a small class, and 31 or more was deemed a large one. Here the results pertaining

TABLE 2

Distribution of Instructional Value-Added

<u>Range</u>	<u>Instructional Value- Added Frequency</u>	<u>Range</u>	<u>Instructional Value- Added Frequency</u>
(0.83) - (0.80)	2	0.40 - 0.49	94
(0.70) - (0.79)	0	0.50 - 0.59	66
(0.60) - (0.69)	6	0.60 - 0.69	16
(0.50) - (0.59)	6	0.70 - 0.79	3
(0.40) - (0.49)	4	0.80 - 0.89	3
(0.30) - (0.39)	32	0.90 - 0.99	5
(0.20) - (0.29)	71	1.00 - 1.09	3
(0.10) - (0.19)	167	1.10 - 1.19	0
0.00 - (0.09)	435	1.20 - 1.29	0
0.01 - 0.09	413	1.30 - 1.39	1
0.10 - 0.19	540	1.40 - 1.49	0
0.20 - 0.29	433	1.50 - 1.59	1
0.30 - 0.39	209		
**(total n = 2,510)			

TABLE 3

Instructional Value-Added and Class Size: Research Question 1

(A) Distribution	<u>Class Size</u>	<u>Frequency</u>	<u>Average Instructional Value-Added</u>
	0 - 5	43	0.151
	6 - 10	208	0.116
	11 - 15	260	0.114
	16 - 20	307	0.115
	21 - 25	217	0.104
	26 - 30	228	0.104
	31 - 35	282	0.142
	36 - 40	229	0.138
	41 - 45	286	0.135
	46 - 50	257	0.125
	51 - 55	130	0.143
	56 - 60	44	0.153
	61 - 65	10	0.139
	>65	9	0.014
	TOTAL	2,510	

(continued)

TABLE 3 (continued)

(B) Statistical RelationshipRegression: Instructional value-added = $\beta_0 + \beta_1 * \text{Class Size} + \varepsilon$

<u>Variable</u>	<u>Predicted Sign</u>	<u>Coefficients</u>
Intercept (b_0)		0.10767 (11.02)***
Class Size (b_1)	-	0.00252 (1.81)*
R-Square		0.0013
Model F Value		3.29
Model p-value		0.0697

*, **, ***: significant at 10, 5, 1%

to class size were also not significant. No support exists for Research Question 1. Instructional value-added does not seem to be an artifact of the size of the student group.

Research Question 2 suggested that instructional value-added may be a product of the inherent characteristics of academic disciplines. With instructional value-added defined as the difference in two ratings, departmental differences would require something more complex than a stronger affective response by students to certain subjects. Table 4 contains descriptive and statistical information pertaining to this question.

TABLE 4

Instructional Value-Added and Academic Departments: Research Question 2**(A) Distribution**

<u>Department</u>	<u>Observations</u>	<u>Average Instructional Value-Added</u>	<u>Average Course Rating</u>	<u>Average Professor Rating</u>
Accountancy	229	0.153	4.113	4.266
Banking and Finance	608	0.105	4.024	4.129
Economics	182	0.157	4.014	4.170
Information Systems	216	0.182	3.941	4.123
Management and Policy	819	0.095	4.130	4.225
Operations	303	0.157	4.170	4.328
Organizational Behavior	153	0.121	4.419	4.541
TOTAL	2,510			

(continued)

TABLE 4 (continued)

(B) Statistical RelationshipRegression: Instructional value-added = $\beta_0 + \beta_1 * \text{College Department} + \varepsilon$

<u>Variable</u>	<u>Predicted Sign</u>	<u>Coefficients</u>
Intercept (b_0)		0.12773 (13.07)***
College Department (b_1)	?	-0.00101 (-0.44)
R-Square		0.0001
Model F Value		0.20
Model p-value		0.6576

*, **, ***: significant at 10, 5, 1%

There is some variation across the seven academic departments. Instructional value-added varies from a high of 0.182 in the Information Systems Department to a low of 0.095 in the Management and Policy Studies Department. Three departments are tightly clustered in the 0.153 to 0.157 range. In no case was total departmental instructional value-added negative on average. Variations may be attributable to differing degrees of emphasis in these departments on quality teaching. Using OLS, the association fails to approximate conventional levels of significance. The statistical results show that instructional value-added is not related to departments ($p > .05$), and Research Question 2 is not supported. Instructional value-added seems individualistic not disciplinary in nature.

The next research question pondered whether instructional value-added might be different in lower and higher level courses. This was based on an expectation that students at various stages in their development would show various degrees of receptivity to the quality of efforts of the faculty. Table 5 shows six levels for these purposes. The university that provided the data uses a fairly conventional course numbering system that roughly parallels the number of years that a student has been matriculated. Thus, freshman and sophomores generally take 100 and 200 level courses. Juniors and seniors generally take 300 level courses. Graduate students take 400, 500 and 600 level courses, with master's level students in the lower ranges. Although these demarcations are not exact, and not mutually exclusive, they are sufficiently logical to provide a measure of matriculation level

The Table 5 results show a somewhat distinct pattern by level. Both average course rating and average instructor rating tends to increase over the undergraduate levels of matriculation (100-300). Both make a solid and close to equal uptick in the movement from the introductory 100 levels courses to the 200 level. A smaller increase of .02 and .03 for the next level (200 to 300) results in the exact rise of 0.12 of the 300 level over the 100 level for both of these metrics. Thus, instructional value-added, as the difference of the two, tends to be fairly constant across all levels of the purely undergraduate courses, varying only .011 across these levels. There appears to be a larger value added difference between courses typically taken by advanced undergraduates (300s) and those

TABLE 5

Instructional Value-Added and the Level of Courses: Research Question 3

(A) Distribution

<u>Course Level</u>	<u>Frequency</u>	<u>Average Instructional Value-Added</u>	<u>Average Course Rating</u>	<u>Average Professor Rating</u>
100 - 199	79	0.198	3.923	4.121
200 - 299	36	0.187	4.021	4.208
300 - 399	226	0.198	4.043	4.241
400 - 499	1,985	0.116	4.082	4.198
500 - 599	136	0.073	4.503	4.576
> 600	48	0.066	4.376	4.442
TOTAL	2,510			

(B) Statistical Relationship

Regression: Instructional value-added = $\beta_0 + \beta_1 * \text{Educational Level} + \epsilon$

<u>Variable</u>	<u>Predicted Sign</u>	<u>Coefficients</u>
Intercept (b_0)		0.26902 (12.51)***
Educational Level (b_1)	?	-0.03744 (-6.87)***
R-Square		0.0185
Model F Value		47.25
Model p-value		<0.0001

*, **, ***: significant at 10, 5, 1%

more routinely populated by graduate students (400s). Whereas average course ratings increased, professor ratings declined. The importance of these observations was tested via an OLS approach that suggested a significant ($p < .01$) statistical relationship. The test was reconsidered, this time excluding doctoral level course numbers (500 and 600). The results were still highly significant, despite considerably lower F-values and t-statistics. Research Question 3 is supported. Instructional value-added has a distinct variation by academic level.

The final exploratory question pertains to the programmatic affiliations of courses. For these purposes, the undergraduate program and several types of graduate programs are differentiated. The graduate side reflects the special nature of doctoral programs and MBA programs. The specialty master programs (from three separate disciplines) may be different from each other but more distinct as a group from the other graduate degrees. This parallels, but does not duplicate, the level of the

course and the academic department previously considered, and is particularly appropriate to the graduate program focus of the university from which the data originates. Table 6 provides descriptive information about instructional value-added levels in the four separate program types. These range from the undergraduate to the doctoral, and separates the generalized masters (MBA) from specialized masters (e.g., Masters of Accountancy).

The results suggest a distinct difference by degree program. Instructional value-added appears to be at its apex in the undergraduate program and at its nadir in the doctoral programs. The results are statistically significant at $p < .01$. Thus, there is evidence of some patterning of factors, such as separate pedagogies or content or student selection. Research Question 4, pertaining to the possibility of programmatic differences, is supported.

TABLE 6

Instructional Value-Added and Degree Programs*: Research Question 4

(A) Distribution

<u>Program</u>	<u>Frequency</u>	<u>Average Instructional Value-Added</u>
Undergraduate	341	0.197
Masters	1,560	0.117
Masters-Specialty	161	0.138
Doctoral Level Programs	448	0.087
TOTAL	2,510	

*Includes Ph.D. offered by five departments and an Executive Doctoral degree

(B) Statistical Relationship

Regression: Instructional value-added = $\beta_0 + \beta_1 * \text{Program Level} + \varepsilon$

<u>Variable</u>	<u>Predicted Sign</u>	<u>Coefficients</u>
Intercept (b_0)		0.17924 (16.86)***
Program Level (b_1)	?	-0.02359 (-5.64)***
R-Square		0.0125
Model F Value		31.78
Model p-value		<0.0001

*, **, ***: significant at 10, 5, 1%

Additional Analyses

In order to provide a closer look at the construct, one discipline was examined with a finer tooth perspective. Perhaps mirroring the accreditation-based belief that the pre-professional training of accountants merits special treatment, courses taught by the accounting department were scrutinized.

The average course and instructor rating for accounting classes (N=210) were 4.10 and 4.26 respectively. Thus, the average instructional value-added was 0.16, or slightly more than the global business school average. Only 40 of the accounting classes (19.0%) produced a negative difference. Accounting classes also had lower variance for both the instructor (0.236) and the course (0.235). The fact that these numbers are much closer than those of the business school is also noteworthy. The average accounting class at this school had 24.5 students during the period. Instructional value-added was not significantly related at $p < .05$ to class size, no matter how class size was measured. Accounting classes existed in several programs including the undergraduate major, the specialized masters' degree and the MBA. Instructional value-added did not significantly vary by program ($p > .10$), although it was directionally higher in the disciplinary-specific Masters of Accountancy program. The examination of course level as a correlate of instructional value-added was also conducted. These results also paralleled the main findings with a significant ($p < .01$) decline of value-added over the range.

To investigate whether the results were an artifact of violations of the normality assumption of linear association, various steps were taken to normalize the data. To counter possible ceiling effects, the data was winsorized at the top end of the distribution by eliminating the top 1% and 5% most highly related courses. This was done under the belief that the top of the Likert scale may have not allowed the full range of value-added. This procedure did not change the statistical conclusions reached in the main body of the findings. Second, outliers at both ends were truncated so that extreme amounts could not skew the results. This gives some recognition to floor effects for classes that might be rated very poorly, opening much more room for instructional value-added. None of these procedures changed the statistical conclusions offered in the previous section.

The differencing technique used in this paper, despite the elegance of its simplicity, may have equated situations that should be distinguished. Specifically, the value-added delivered by a faculty person who delivers a poorly-received class may be qualitatively different, than that produced by an instructor rating that exceeds a well-rated class by the same magnitude. To consider this, the data was re-examined after a slight (10%) and a more extreme (20%) weighting process such that the instructors with more highly rated courses were credited with higher weights for the spread between the instructor and course rating. This procedure draws justification from the fact that the course rating may be influenced by the teacher. If part of the instructor's task is to inculcate an appreciation for the material of the course, those more successful at this may have suppressed their true value-added as a teacher, as measured here. The results show that this adjustment is not capable of altering the statistical conclusions related to any of the hypotheses. The rank ordering of the departments reported in Table 4 were largely unchanged over both weights studied, with only Organizational Behavior and Information Systems repositioning in a downward and upward direction, respectively.

In order to study the possible instability of the benchmark course rating, a new baseline from which value-added could be measured was used. Here, the average course evaluation for each course was computed across all instructors and across all years. This alters the conceptualization of value-added from that transcendence over the material of that class to the transcendence over the average

perception of this class's material by all students who have taken it from any instructor. This avoids the potential tendency for the value-added contributions of teachers to be confounded with how the course is perceived. Unlike the weighting procedure described above (which increased instructional value-added) the use of this new benchmark to measure instructional value-added does not have the effect of increasing the average of this metric (still equal to 0.12). It does produce a standard deviation nearly twice as large, however. The only change observed in the re-testing of the hypothesis is to make class size clearly not significant. The original results had been that class size was marginally significant ($p < .10$). In other hypotheses, despite lower F values for the regressions, the previously reported results were replicated.

Discussion

The results demonstrate that instructional value-added is a reasonably stable variable that can be used to evaluate college-level teaching. The construct does not seem to reflect the intimacy of the setting, insofar as it is not related to class size. The evidence also suggests that it does not put some disciplines at an unfair advantage. To that extent, instructional value-added appears to provide information not duplicated by the absolute position on the Likert scales that have been combined in its production.

The results suggest, however, that instructional value-added is, to a limited extent, in the eye of the beholder. More advanced students are increasingly likely to equate their reaction to the course material and their instructor. Individuals in the earlier stages of their studies are more likely to make these metrics separate. Thus, instructional value-added is a result of student maturation, with younger students more dependent upon the instructor's capabilities. This was approached in the empirical work via the importance of the level of the course and the separation of different programs of study.

Some have argued that the value of student evaluations is so small that they should not be administered (e.g., Wallace and Wallace, 1998). This study weighs in against this position. The problem with the use of evaluations centers around the literal belief in their absolute position on the scale. In other words, interpretation of the results must be put into a context. The simple differencing that lies at the heart of this study requires those who use these instruments, or are subject to them, to move away from this traditional orientation. Instructional value-added forces the question – is it the material or is it the teacher? Such more focused inquires might bring the use of teaching evaluations back to its original intention to serve as formative feedback device to allow the faculty person to change that which can be improved.

The results of this study can be interpreted in the light of recent forces that have sought to commodify the teaching of college material. Various forms of technologically mediated instruction invite us to focus on the content, rather than upon the delivery, of courses. While course content is important, we should not neglect the instructor's separate contribution. This paper argues that we need to do more to identify and measure what value the faculty bring to education. If instructional value-added would be either lost or greatly compromised without the live and spontaneous presence of a faculty person engaging with students in the process of learning, an opportunity cost of considerable potential consequence needs to be more broadly acknowledged.

In a similar vein, this study suggests that the experience of education is not completely driven by content. There is a value-added component whereby a faculty person makes the content come alive for the student. The instructor's contribution has been found to be very important to the students' overall educational experience (Marsh, 1984). In other words, instructional value-added

as a variable heightens our appreciation for the process of post-secondary education. Without this human touch, that combines the role modeling of the life of the mind with the value of a specific domain, a college degree would be no different than a library card and an internet connection.

This study also could be seen as an implicit argument for the refinement of how teaching is rewarded in the academy. Although there is a strong case for considering that quality teaching can never be reduced to the numbers of an evaluative scale, metrics of this sort will continue to exert considerable enough interest or fixation to merit their best use. The identification of a reasonably stable instructional value-added component in the ratings requires that we confront what it is that we value about teaching. This paper essentially argues that an absolute and literal reading of these numerical summaries is not as valuable as a more subtle attempt to isolate the teaching component. What this paper calls instructional value-added is a first attempt to move toward this valuable objective.

Limitations

Many studies have attempted to convince us that the teaching act is very complex and therefore in need of the combination of multiple sources to evaluate information (e.g., Calderon and Green, 1997). Other studies have argued that the numbers produced by teaching evaluations are biased reflections of the context of their production (e.g., Ketler et al. 2000). Both of these traditions illustrate potential sources of limitations of the current study. In fact, the ratings cannot capture everything that should be valued about teaching and therefore are less than we would want in a perfect world. On the other hand, the ratings may also include influences that should be extraneous to our ideas about quality instruction. This would suggest that we have a measure that is over-enriched. To the extent that both have elements of truth, they form opposite and somewhat offsetting limitations of the present work. However, as long as the academy is also very interested in conducting research, and provide services to many constituents, there will be forces that will necessitate the use of teaching metrics that are deemed “good enough.” If this is the case, we should learn to use what we have more intelligently.

Critics of studies who use student-prepared evaluations often say that these instruments have little or no bearing upon actual learning. This seems to be the prevailing belief, even among those who are charged with the oversight of this information (e.g. Calderon et al. 1996). Those who insist upon stronger measures of learning include those who are struggling to find completely different frameworks (e.g., Booth et al. 1999) and those who want to link to performance standards outside the academy (e.g., Schick, 1998). The connection between teaching evaluations and student learning is problematic. Even in the presence of institutional agreement on objectives and measurements, inevitable subjectivity in interpretation persists. Some progress has been made on finding associations between specific instructor behaviors (captured on student evaluation instruments) and student achievement (e.g., Feldman, 1989). *Ceteris paribus*, most would agree, however, that teacher effectiveness is an important motivating and facilitating factor in student learning. Therefore, ways to measure the incremental contribution of faculty will remain necessary.

The approach suggested by this paper should be seen as a supplement to conventional faculty evaluation practice. It cannot replace concern with the absolute level of student evaluation. Faculty with higher teaching and course evaluations are to be praised more than those with lower evaluations, notwithstanding the value-added calculation advocated by this paper. Nonetheless, instructional value-added provides heightened interpretation possibilities for the absolute results.

The data presented in this paper come from a single school, and therefore generalization to the entirety of the business academy might not be appropriate. An extrapolation is especially problematic as applied to the specific business school disciplines, due to the smaller number of people involved. Because any academic department combines elements of the discipline, the historical singularity of the school and the personalities of the human beings, generalization should be cautious.

Implications

As an exploratory study, this paper opens many new lines of research. For example, student perception of quality teaching may be related to the pedagogical choices made by the instructor (see Marsh, 1987). In addition to small methodological choices (such as the deployment of cases), and more integrative efforts to merge content with skills, pedagogical innovation may be a double-edged sword as it is received by students (see Reynolds, 1999). Working within established methods, even as they nourish stereotypes (see Mladenovic, 2000), may provide psychological comfort for self-selecting students. These conservative tendencies of students might disrupt expectations and depress course evaluations. However, to the extent that students credit instructors for their unique effort, the result might be a better ability to identify the value-added of innovative pedagogy. Particularly resourceful teachers can be detected by higher instructional value-added scores, more than through the absolute evaluative levels that they achieve. However, much more research is needed to explicate the relationship between teacher creativity and student response.

The identification and preliminary evaluation of instructional value-added invites the question of institutional reaction. This issue may depend upon the centrality of teaching to the mission of the school (Raghunandan et al. 1999). Schools will have different ways of translating the instructional value-added factor into quality increments. For example, instructional value-added could be “spun” as evidence of quality itself, or decried as indicative of showmanship or entertainment divorced from quality. Perhaps, the instructional value-added factor will hasten a search for an alternative classroom assessment that is arguably more connected to student learning (e.g., Cottell and Harwood, 1998). Watching these developments will be provocative to an education-oriented faculty. Short of that, instructional value-added may be correlated to some of the many highly specific areas that teaching evaluations already measure, such as those that capture the instructor’s “emotional intelligence” (Goleman et al. 2002). Since no singular criteria of effective teaching exists (Marsh, 1984), research in this area still grapples with foundational matters.

The construction of instructional value-added depends in no small way upon the personality of the instructor. By itself, that contingency could keep many researchers busy since the literature has no shortage of ways to approach personality. More dynamically, the interaction between personality and methods should be explored. A faculty person should have sufficient self-insight so as to select the most appropriate instructional methods for his or her personality. Gender differences may also have to be considered, especially if the sexes have different ways of knowing (Gallos, 1993). Skeptics will also have to be convinced that instructional value-added cannot be reduced to the possession of a sense of humor, an attribute believed to be important to the teaching art (see Tomkovich, 2004). Instructional value-added may be steeped in the instructor’s “soft skills” as much as in their ability to communicate the substance of their academic subjects. Collectively, this work should expand our appreciation for the art of teaching.

The confirmation of the Research Questions that explored variation by level and program point toward the importance of the learner. The instructional value-added of the teacher does not

seem to be a constant but instead is the product of the general learning process that occurs during post-secondary education. Younger students may rate their introductory courses lower because they fail to fully appreciate their relevance. Students in the earlier stages of coursework are more impressed by their professors and may therefore overestimate their contribution. Part of the growing sophistication of the student may involve the ability to peak behind the wizard's curtain. Since instructional value-added does not operate in a social psychological vacuum, further research is necessary to study the extent of these influences over the course of the student's post-secondary career.

REFERENCES

- Ainger, D., and F. Thum. 1986. On Student Evaluation of Teaching Ability. *Journal of Economic Education* (Vol. 17) 243-265.
- Atkinson, P., and S. Delamont. 1990. Writing About Teaching: How British and American Ethnographic Texts Describe Teachers and Teaching. *Teaching and Teacher Education* (Vol. 6) 111-127.
- Bailey, A., and W. Bentz. 1991. Accounting Accreditation: Change and Transition. *Issues in Accounting Education* (Vol. 6) 168-177.
- Bailey, C., S. Gupta, and R. Schrader. 2000. Do Student Judgment Models of Instructor Effectiveness Differ by Course Content or Individual Instruction? *Journal of Accounting Education* (Vol. 18) 15-34.
- Bell, T., T. Frecka, and I. Solomon. 1993. The Relationship Between Research Productivity and Teaching Effectiveness: Empirical Evidence for Accounting Educators. *Accounting Horizons* (Vol. 7) 33-49.
- Benke, R., and B. Roof. 1990. Scholarly Productivity and Teaching Effectiveness. *Management Accounting* (Vol. 72, no. 6) 54-55.
- Booth, P., P. Luckett, and R. Mladenovic. 1999. The Quality of Learning in Accounting Education, The Impact of Approaches to Learning on Academic Performance. *Accounting Education* (Vol. 8) 277-300.
- Briscoe, N., G. Glezen, and W. Letzkus. 1998. The Association of Accounting Course Content Groupings and Student Evaluation. *Accounting Educators' Journal* (Vol. 8) 14-26.
- Buckmaster, N., and R. Craig. 2000. Popular Television Formats, the Student-as-Consumer Metaphor, Acculturation and Critical Engagement in the Teaching of Accounting. *Accounting Education* (Vol. 9) 371-387.
- Calderon, T., A. Gabbin, and B. Green. 1996. Summary of Promoting and Evaluating Effective Teaching. *Journal of Accounting Education* (Vol. 14) 367-383.
- _____, and B. Green. 1997. Use of Multiple Information Types in Assessing Accounting Faculty Teaching Performance. *Journal of Accounting Education* (Vol. 15) 221-239.
- _____, _____, and B. Reider. 1997. Perceptions and Use of Student Evaluations by Heads of Accounting Departments. *Accounting Educators' Journal* (Vol. 9) 1-27.
- Cares, R., and R. Blackburn. 1978. Faculty Self-Actualization: Factors Affecting Career Success. *Research in Higher Education* (Vol. 9) 123-136.
- Carruth, P., and A. Carruth. 1997. Evaluating Teacher Effectiveness. Paper presented at the Midwest Business Administration Meetings, Chicago, IL.
- Centra, J. 1983. Research Productivity and Teaching Effectiveness. *Research in Higher Education* (Vol. 18) 379-389.

- Chen, Y., and L. Hoshower. 1998. Assessing Student Motivation to Participate in Teaching Evaluation: An Application of Expectancy Theory. *Issues in Accounting Education* (Vol. 13) 531-550.
- Clark, B. 1987. *The Academic Life* (Princeton: Carnegie Foundation).
- Cottell, P., and E. Harwood. 1998. Using Classroom Assessment Techniques to Improve Student Learning in Accounting Classes. *Issues in Accounting Education* (Vol. 13) 551-564.
- D'Onofrio, M., M. Slama, and A. Tashchian. 1988. Faculty Evaluation in Colleges of Business. *Journal of Marketing Education* (Vol. 8) 21-28.
- Dopuch, N. 1989. Integrating Research and Teaching. *Issues in Accounting Education* (Vol. 4) 1-10.
- Feldman, K. 1987. Research Productivity and Scholarly Accomplishment of College Teachers as Related to Their Instructional Effectiveness. *Research in Higher Education* (Vol. 26) 227-291.
- _____. 1989. The Association between Student Ratings of Specific Instruction Dimensions and Student Achievement. *Research in Higher Education* (Vol. 30) 583-645.
- Gallos, J. 1993. Woman's Experiences and Ways of Knowing: Implications for Teaching and Learning in the Organizational Behavior Classroom. *Journal of Medical Education* (Vol. 17) 1-26.
- Goleman, D., R. Boyatzis, and A. McKee. 2002. *Primal Leadership*. (Boston: Harvard Business Press).
- Green, B., T. Calderon, A. Gabbin, and J. Habegger. 1999a. Perspectives on Implementing a Framework for Effective Teaching. *Journal of Accounting Education* (Vol. 17) 71-98.
- _____, B. Reider, and T. Calderon. 1999b. Biases in Student Evaluation of Teaching. Paper presented at the South East Regional Meeting of the American Accounting Association.
- Harwood, E. 1999. Student Perception of the Effects of Classroom Assessment Techniques. *Journal of Accounting Education* (Vol. 17) 51-70.
- Hattie, J., and H. Marsh. 1996. The Relationship between Research and Teaching: A Meta-Analysis. *Review of Educational Research* 66(4): 507-542.
- Henke, D. 1998. How Instructors of Accounting in Higher Education Spend Their Time, Earn Their Compensation, and Chose Their Professional Organization. Paper presented at the Midwest Business Administration Meetings, Chicago, IL.
- Jackson, S. 1996. Dealing with the Overenriched Work Life. *Rhythms of Academic Life* (P. Frost and L. Cumming eds.) (Thousand Oaks CA: Sage Publications). 351-355.
- Kalbers, L., and G. Weinstein. 1999. Student Performance in Introductory Accounting: A Multi-Sample, Multi-Model Analysis. Paper presented at the American Accounting Association Annual Meeting, San Diego, CA.
- Ketler, K., J. Walstrom, R. Schelhavy, N. Maslow, and E. Marlow. 2000. Teaching Survey of Accredited Collegiate Schools of Business: A Comparison of Departments, unpublished paper.
- Langbein, L. 1994. The Validity of Student Evaluations of Teaching. *Political Science and Politics* (Vol. 27) 545-553.
- Lewis, L. 1996. *Marginal Worth: Teaching and the Academic Labor Market*. (New Brunswick, NJ: Transaction Publishers).
- _____. 1968. Student Images of Professors. *Educational Forum* (Vol. 32) 185-190.
- Light, R. 2001. *Making the Most of College*. (Cambridge MA: Harvard University Press).
- Marsh, H. 1984. Student's Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Utility. *Journal of Educational Psychology* (Vol.7 No.5): 707-754.

- _____. 1987. Students' Evaluation of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. *International Journal of Educational Research*. (Vol.11) 253-288.
- _____, and M. Bailey. 1993. Multidimensional Students' Evaluation of Teaching Effectiveness: A Profile Analysis. *Journal of Higher Education* (Vol.64) 346-411.
- McKeachie, W. 1980. Class Size, Large Classes and Multiple Sections. *Academe* (Vol. 66) 24-27.
- Meyer, M., and P. Titard. 2000. Those Who Can Teach. *Journal of Accountancy* (Vol. 184) 49-58.
- Mladenovic, R. 2000. An Investigation into Ways of Challenging Introductory Accounting Students' Negative Perceptions of Accounting. *Accounting Education* (Vol. 2) 135-155.
- Nisbett, R., and T. Wilson. 1977. Telling More than We Can Know. *Psychological Review* (Vol. 86) 231-259.
- Osborne, E., and J. Lukshin. 2000. What Do Accounting Students Really Want in the Classroom? Presented at the annual meeting of the Midwest Business Administration Association, Chicago.
- Palmer, P. 1998. *The Courage to Teach: Exploring the Inner Landscape of a Teacher's Life*. (San Francisco: Jossey-Bass).
- Raghunandan, K., W. Read, and D. Rama. 1999. The Importance of Teaching and Reliance on Student Evaluations of Teaching. Presented at the American Accounting Association Annual Meeting, San Diego.
- Read, W., D. Rama, and K.Raghunandan. 1998. Are Publication Requirements for Accounting Faculty Promotions Still Increasing? *Issues in Accounting Education* (Vol. 13) 327-340.
- Reckers, P. 1996. Know Thy Customer. *Journal of Accounting Education* (Vol. 14) 179-185.
- Reynolds, M. 1999. Grasping the Nettle: Possibilities and Pitfalls of a Critical Management Pedagogy. *British Journal of Management* (Vol. 9) 171-184.
- Root, L. 1987. Faculty Evaluation: Reliability of Peer Assessments of Research, Teaching and Service. *Research in Higher Education* (Vol. 26) 71-84.
- Schick, A. 1998. Should Undergraduate Education in Accounting be Evaluated in Part Based on Graduates' Performance on the CPA Examinations? *Issues in Accounting Education* (Vol. 13) 417-420.
- Seldin, P. 1984. Faculty Evaluations: Surveying Policy and Practices. *Changes* (Vol. 16) 28-33.
- Stevens, K., and W. Stevens 1992. Evidence on the Extent of Training in Teaching and Educational Research among Accounting Faculty. *Journal of Accounting Education* (Vol. 10) 271-283.
- Street, D., C. Baril, and R. Benke. 1993. Research, Teaching and Service in Promotion and Tenure Decisions of Accounting Faculty. *Journal of Accounting Education* (Vol. 11) 112-134.
- Tomkovich, C. 2004. Ten Anchor Points for Teaching Principles of Marketing. *Journal of Marketing Education* (Vol. 26, No. 2) 109-115.
- Wachtel, H. 1998. Student Evaluation of College Teaching Effectiveness: A Brief Review. *Assessment and Evaluation in Higher Education* (Vol. 23) 191-212.
- Wallace, J., and W. Wallace. 1998. Why the Costs of Student Evaluations Have Long Exceeded Their Value. *Issues in Accounting Education* (Vol. 13) 443-448.
- Wisdom, B., and H. Teer. 1990. Role Perceptions of the Accounting Educator: Presently and Ideally. *Accounting Educators' Journal* (Vol. 2) 51-61.
- Wright, P., R. Whittington, and G. Whittenburg. 1984. Student Rating of Teaching Effectiveness: What the Research Reveals. *Journal of Accounting Education* (Vol. 2) 5-30.

APPENDIX

Course Evaluation Questions Considered by the Students for the Empirical Portions of this Research

1. What is your overall rating of the course?
2. What is your overall rating of the instructor?