Using SERVQUAL to Measure the Quality of the Classroom Experience

Michael Stodnick† and Pamela Rogers
Department of Management, College of Business Administration, University of North Texas, 1167 Union Circle Denton, TX 76201, e-mail: stodnicm@unt.edu, rogersp@unt.edu

ABSTRACT

Over the last three decades, higher education institutions have found themselves using vernacular that was once chiefly found in business disciplines, such as value-added and competitive advantage. With the rising costs of tuition, newer-generation students are seeing themselves more and more as customers and universities are beginning to adopt customer-centric strategies and missions. However, even with this paradigm shift, little research has been done to extend traditional service management concepts to educational settings. This research attempts to bridge this gap by applying the SERVQUAL scale, a well-validated and widely used service operations construct, to the classroom environment. The findings show that the SERVQUAL scale exhibits both reliability and convergent and divergent validity; in fact, in these regards, it outperforms traditional student assessment scales. Moreover, the scale can explain significant amounts of variances in student-related outcome variables such as satisfaction and learning. This innovative approach to measuring classroom service quality does indeed show that students can be viewed as customers and has far-reaching implications to all stakeholders in the delivery of higher education.

Subject Areas: Service Quality, SERVQUAL, Student Evaluations, Student Learning, and Student Satisfaction.

INTRODUCTION

Over the last three decades, higher education institutions have experienced dramatic shifts in both their funding formulas and student populations. Creating a competitive advantage, once a concept largely foreign to higher education, has become a driving force (Oldfield & Baron, 2000). The myriad of stakeholders involved in or influenced by higher education are now seeking evidence of institutions’ effectiveness in achieving educational goals. Although consensus among these stakeholders as to the definition of quality education may vary by segment, the stakeholders are of the same mindset in calling for indicators that capture performance of all those involved in executing and improving the delivery of higher education (Nedwek & Neal, 1994).

†Corresponding author.
The intensified competition within higher education mirrors that found within the service sector in general. The response of many firms to the heightened call for enhanced quality was to implement continuous improvement programs such as total quality management and/or Six Sigma. A key tenet to these philosophies is that organizations should continually assess customer perceptions of service quality. Only when data are collected and analyzed can real improvements be made (Jensen & Artz, 2005). Universities are giving serious consideration to the issue of service quality assessment for a multitude of reasons, arguably the two most important of which are: students report that word-of-mouth recommendations play a large role in their decision to choose a university and both university quality assurance and independent assessment evaluators place heavy emphasis on the student experience as one of their assessment criteria (Cuthbert, 1996). The underlying theory is that institutions that continually improve service quality and delivery are more likely to generate high levels of customer satisfaction, resulting in both increased customer loyalty (namely, a higher retention of the current student population), and decreased costs of attracting new students (through positive word of mouth from the students and higher independent ratings).

In this study, we focus squarely on one portion of what Petruzzellis, D’Uggento, and Romanazzi (2006) call the “total student experience”—that is, the quality of the classroom encounter. The use of student ratings to provide feedback about the quality of instruction developed out of protests in the late 1960s from students who increasingly saw themselves as customers (Centra, 1993). Since that time, a vast number of studies, including several meta-analyses, have shown that the use of student ratings is both a reliable and valid measurement of teaching quality. A review of assessment literature conducted by Brightman et al. (1993) concluded that ratings provide normative data that can be used as a mechanism for teaching improvement.

Recently, this customer-centric approach of service quality has gained momentum in educational literature as the increasing cost of education has created a new generation of students with greater customer awareness than ever before. As Oldfield and Baron (2000) pointed out, the “interaction between customer and service organization lies at the heart of the service delivery.” Employees who deliver the service, in this case the instructor, are of key importance to both the customers they serve, the students, and the employer they represent, the university. In some regards, the employee (instructor) may be the most visible route by which the employer (the university) can distinguish itself.

The principal instrument used in service management and marketing literature to measure service quality is the SERVQUAL scale. However, even as higher education continues to strive toward customer-oriented strategies, very little work has been done to combine education literature with service management and marketing research. This research bridges this gap by applying the SERVQUAL scale within a classroom setting. Can SERVQUAL, a valid and reliable customer-centric scale used to measure the quality of service delivery in environments as diverse as retail and business consulting, be used to measure and thus ultimately improve the quality of service delivery in higher education? In other words, can this well-validated scale be innovatively applied to measure student perceptions of classroom delivery? This question is of paramount importance to all stakeholders in higher education. Better measures of the customers’ voices through their assessment of
service quality may ultimately lead to improved educational experience (student), increased professional development (instructor), higher university ranking (university itself), better-qualified graduates (community), and so on.

LITERATURE REVIEW

Full reviews of SERVQUAL and student evaluation literature are well beyond the scope of this article. Instead, after a brief summary of what the SERVQUAL scale is and its inception, we will focus on applications of SERVQUAL in higher education. The conceptual underpinnings of the SERVQUAL model were first published in 1985 (Parasuraman, Zeithaml, & Berry, 1985). In that research, the authors focused their discussion of service quality on what Gronroos (1984) labeled “functional quality,” or the expressive performance of a service. They argued that there are 10 distinct dimensions to service quality. However, 3 years later when empirically deriving a service quality definition, the list of 10 was reduced to 5; the 5 dimensions and the descriptions the authors give are listed below (Parasuraman, Zeithaml, & Berry, 1988).

- **Tangible**—physical facilities, equipment, and appearance of personnel;
- **Reliability**—ability to perform the promised service dependably and accurately;
- **Responsiveness**—willingness to help customers and provide prompt service;
- **Assurance**—knowledge and courtesy of employees and their ability to inspire trust and confidence; and
- **Empathy**—caring, individualized attention the firm provides its customers.

Over the last 20 years, authors have used the SERVQUAL scale to measure service quality in a wide selection of industries with varying success. The primary emphasis of these studies has been to test whether SERVQUAL is an appropriate measure of service quality in varying contexts and to determine the antecedents and consequences of delivering superior service quality. For a full review of the current and future states of SERVQUAL research, see Parasuraman and Zeithaml (2002).

We have identified five studies that have applied SERVQUAL within a university environment. Cuthbert (1996) pioneered this stream of research by examining the applicability of the SERVQUAL scale to measure student perceptions of university-level service quality. The author began by testing the reliability of the five SERVQUAL dimensions and found very weak results: Cronbach alpha scores ranging between .01 and .52. Because of these lower-than-expected scores, the SERVQUAL items were subjected to exploratory factor analysis. Seven factors formed and, as the author pointed out, these new factors little resembled the original five factors. The author concluded from these results that using the SERVQUAL scale to measure university service quality seems inappropriate. No analysis was performed to determine whether any of the items in SERVQUAL can be used to predict student satisfaction or any similar dependent variable.

Several studies extended Cuthbert’s (1996) initial work in this area. Oldfield and Baron (2000) replicated the Cuthbert (1996) study 4 years later, using
SERVQUAL to measure student perceptions of business and management faculty. Through an application of exploratory factor analysis, the researchers found that three factors emerge: requisite, essential items that allow students to fulfill their study obligations; acceptable, items that are preferable rather than essential to student development; and functional, items outside the control of the instructor and primarily derived from university rules. The authors did not test the link between these factors and student outcome measures. Sahney, Banwet, and Karunes (2004) explored the possibility of using SERVQUAL to measure student perceptions of service quality in higher education in India. Their factor analysis suggested that the SERVQUAL items were actually unidimensional. Although the authors showed how the SERVQUAL items can be used in quality function deployment to improve the university’s services, no analysis was done to indicate whether any SERVQUAL items can be predictive in nature.

Two other studies used SERVQUAL in a more focused way. Hughey, Chawla, and Khan (2003) employed the SERVQUAL instrument to measure service quality of computer labs, carrying out two studies separated by a 2-year interval. In both studies, the SERVQUAL items load onto three factors: staff, services, and professionalism. The authors conducted several tests to investigate whether gender, academic standing, and time spent in labs influence a student’s perception of service quality. They found that female students tended to rate the university more highly on the services and professionalism constructs than did their male counterparts. The only other significant result was that juniors rated staff higher than seniors. O’Neill (2003) studied the application of SERVQUAL in a university orientation setting. Students were asked to assess the quality of the orientation process immediately after orientation, time $t$. One month later, time $t + 1$, the same students were asked to reflect back on the orientation process and fill out another survey. The author’s analysis showed that SERVQUAL was unidimensional at time $t$, but three-dimensional at time $t + 1$. Neither of these two studies investigated whether SERVQUAL was a predictor of student satisfaction or any other dependent variable.

This article further develops the application of SERVQUAL in an academic setting and makes several key contributions. To begin with, this study is the first to use the SERVQUAL model within a particular class, which could be considered a narrower service encounter than previous studies have used. We make this choice for several reasons. By selecting one class, we will not aggregate data to the extent that much of the previous SERVQUAL education research has done. Previous research that applied SERVQUAL to measure student perceptions of the overall university may not be capturing all the important variance in these student perceptions. For example, if a student has one excellent instructor and one poor instructor and is asked about the quality of instruction at the university, they might answer “average.” While technically this is correct, it would be hard to use their response to make specific changes and improve service quality. It would not be fair or a good use of resources (e.g., instructor time) for the university to tell the excellent instructor to improve because overall instruction is “average.” Likewise, if the poor instructor reviews the results of the study, he or she may feel that he or she has no need for improvement. In contrast, our application will target the instructor delivering the service, thereby giving the instructor detailed, actionable information. The instructor can glean information and improvement areas not only
when SERVQUAL is applied to a single class they teach (giving them specific information for that particular class) but also when they aggregate their scores across all their classes (giving them more general information about their instruction techniques).

A second significant contribution that this study makes is the use of a more comprehensive methodology in answering the question of whether or not it is appropriate to use SERVQUAL to measure quality of instruction. Simply testing the dimensionality of SERVQUAL, as previous studies have done, is only a first step. Another vital part of the equation is comparing SERVQUAL to other student evaluation scales. When compared to other scales, SERVQUAL’s reliability and its ability to predict other student measures and outcomes should be similar to existing scales. The scale we use for comparison is Brightman et al. (1993), hereto referred to as the Brightman scale. We have chosen the Brightman scale because of its widespread use in pedagogy literature.

The underpinnings of the Brightman scale can be found in three places. First, many of the items used in the scale are derived from Berkeley Student Description of Teaching instrument developed by Davis, Wood, and Wilson (Wilson, 1986). The Wilson (1986) article that describes the survey has been cited more than 45 times; examples include Boex (2000) and Pietier, Drago, and Schibrowsky (2003). Furthermore, the Brightman scale relies heavily on two meta-analyses of pedagogy literature—Centra (1987) and Feldman (1989). Essentially, both of these studies put together a list of the most common items used in student assessments and reported which items were best in predicting various outcome measures. These listings heavily influence the items used in the Brightman scale. Brightman et al. (1993) applied factor analysis to a 34-item survey resulting in a six-dimensional scale: organization and clarity, communication ability, grading and assignments, interaction with students, intellectual and scholarly, and student motivation. This instrument is the one currently used at the university where this research is conducted.

A third major contribution of this study is testing whether the SERVQUAL scale has predictive ability in a classroom setting. For a service quality scale to be meaningful and useful, it must not only reliably describe customer perceptions of quality but also have a significant relationship with other customer measures in order to be actionable. An abundance of service quality literature exists in this area; a small selection of which is reviewed later. Outside of the classroom setting, service quality has been shown to have significant positive impact on customer satisfaction (Cronin & Taylor, 1992; Marley, Collier, & Goldstein, 2004; Voss, Parasurama, & Grewal, 1998), customer loyalty (Aldagian & Buttle, 2002; Lee, Lee, & Yoo, 2000; McDougal & Levesque, 2000), and various profitability and market-related performance measures (Kamurka, Mittal, de Rosa, & Mazzon, 2002; Silvestro & Cross, 2000; Zeithaml, Berry, & Parasuraman, 1996). Within the classroom setting, Jensen and Artz (2005) have shown that the positive relationship between service quality and satisfaction with both instructor and course holds true. Pedagogy literature also indicates that students who feel that they receive high-quality instruction report higher learning and development levels than students who do not perceive quality instruction (Cabrera, Colbeck, & Terenzini, 2001). To be valid, the instrument used to measure service quality in a classroom environment must support this link. As Gibbs (1995) points out, students often prefer what is actually
detrimental to their long-term development. As such, a survey must contain items that are linked to student performance and learning and not just questions that amount to nothing more than a popularity contest.

Finally, there is a concern in educational literature that student evaluations are biased by grade expectations (see Eiszler, 2002, and Marks & O’Connell, 2003, for brief reviews). If SERVQUAL is indeed applicable in a classroom setting, it should be free from this bias. This study will be the first to investigate whether this bias exists.

RESEARCH OBJECTIVES

The specific objectives of this study are

- to investigate whether the SERVQUAL scale will be reliable and valid in a university classroom setting;
- to determine whether the SERVQUAL scale exhibits predictive validity by testing its relationship with student satisfaction and learning measures;
- to compare the reliability and validity of the SERVQUAL scale to that of another well-established student evaluation scale, Brightman et al. (1993); and
- to explore whether the SERVQUAL scale is free of grade expectation bias.

HYPOTHESES REVIEW

Four sets of hypotheses were tested in this study. They are (alternative hypotheses are omitted for brevity):

- **H1a**: Student evaluations of service quality, as measured by the Brightman et al. (1993) scale, are positively associated with student satisfaction with the course.
- **H1b**: Student evaluations of service quality, as measured by the SERVQUAL scale, are positively associated with student satisfaction with the course.
- **H2a**: Student evaluations of service quality, as measured by the Brightman et al. (1993) scale, are positively associated with student satisfaction with the instructor.
- **H2b**: Student evaluations of service quality, as measured by the SERVQUAL scale, are positively associated with student satisfaction with the instructor.
- **H3a**: Student evaluations of service quality, as measured by the Brightman et al. (1993) scale, are positively associated with student perceptions of learning.
- **H3b**: Student evaluations of service quality, as measured by the SERVQUAL scale, are positively associated with student perceptions of learning.
- **H4a**: There will be no significant correlation between a student’s expected grade and their evaluation of service quality, as measured by the Brightman et al. (1993) scale.
- **H4b**: There will be no significant correlation between a student’s expected grade and their evaluation of service quality, as measured by the SERVQUAL scale.
METHODOLOGY

Sample and Data Collection

The sample for this research consisted of six undergraduate Operations Management courses at a large southwestern university. Four sections of Introduction to Operations Management, one section of Purchasing, and one section of Production Planning and Control were surveyed. Although individual responses were anonymous, descriptive statistics of the students enrolled in the courses were calculated. The total population size was 264, of which 58% were male and 42% female. Eighty-eight percent of the population were from the business school, 7% from the engineering school, 2% from arts and sciences, and 3% divided among the other schools and/or undecided. Ninety-eight percent of the population were undergraduate students composed of 74% seniors, 23% juniors, and 1% sophomores, while 2% of the population were postgraduate students.

Survey Instrument

This research used an anonymous online survey to collect the data. Each student in the six classes was asked to voluntarily fill out the survey at the end of the semester. The questions used on the survey were derived from previous studies. The 34 questions used to measure the six instructor rating constructs of the Brightman scale were taken verbatim from the Brightman et al. (1993) study. All of these items were measured on a 5-point Likert scale. The 19 questions used to measure the five SERVQUAL dimensions were adapted from the study by Parasuraman, Zeithaml, and Berry (1991). As suggested by the authors, the wording was changed to fit the classroom environment (see Appendix A for the 19 SERVQUAL items). Following the advice of Oldfield and Barron (2000), who argued that perception-only scores should be used when there is a long time delay between assessing expectation and performance, the survey used a perceptions-only scale, as approximately 3 months intervene between the forming of expectations at the beginning of the semester and the rating of performance at the end. The perceptions-only scale, sometimes called SERVPERF, has been validated in a number of research settings (Cronin & Taylor, 1992, 1994; Lee et al., 2000). The items measuring the five constructs used 5-point Likert scales. One question was used to measure each of the following dependent variables: overall student satisfaction with the course, overall student satisfaction with the instructor, amount the student learned throughout the course (Cabrera et al., 2001), and expected grade (Eiszler, 2002). The first three questions used a scale of 1 (lowest) to 10 (highest), the expected grade was measured on a scale ranging from 0 (F) to 4 (A). The total sample size derived from the online survey was 198, which yields a response rate of 75%.

Reliability and Validity—Brightman Scale

Because both scales used in this study have been well established in previous literature, a confirmatory factor analysis approach was used in scale development. We began by testing the reliability of the six individual constructs. Reliability assessment was done using two different methodologies: corrected item to total correlation (CITC) (Kerlinger, 1986) and Cronbach alpha (Nunnally, 1978). The CITC method posits that each item within a construct should be highly correlated
with the construct itself. Kerlinger (1986) recommends that every item within the scale should have a CITC value that exceeds .4. The lowest CITC value for the 34 questions (on their respective six constructs) was .543 for the fifth item in the grading and assignments construct; in fact, that was the only value below .6. The Cronbach alpha values, measures of internal consistency, are presented in Table 1. The lowest value, .758, is well above Nunnally’s suggested cutoff of .7. All others are .86 or above.

After assessing the scales’ reliabilities, we turned to an exploration of both convergent and divergent validity. Convergent validity is the extent to which indicators are associated with each other and represent a single concept. Divergent validity is the degree to which a construct and its indicators differ from other constructs and their indicators. We tested for convergent validity by examining: the structure of the eigenvalues (a factor should only have one eigenvalue over 1.0); percent of variance explained (the items in the factor should explain at least 40% of the variance in the factor); and factor loadings of each construct (all factor loadings should exceed .4) (Ahire & Deveraj, 2001). Table 1 contains the results. Each factor appears to converge toward unidimensionality. We assessed divergent validity for each construct by calculating the Cronbach alpha minus average interscale correlation (AVISC) value. The difference between the two should be substantially greater than zero. Although there is no statistical test of significance, difference values of .3 and .4 have been used in the past (McDougal & Levesque, 2000; Petrick, 2002; Spreng & MacKoy, 1996). The last column in Table 1 summarizes the findings. All six of the scores presented are very low; indicating that the scales may not be measuring six distinct concepts, and the scales may be suffering from multicollinearity.

Because the main analysis will use multiple regression, a technique highly sensitive to multicollinearity, we decided to explore this issue further. For each of the six constructs variance inflation factor (VIF) scores were calculated. In this case, the VIF scores tell how well the sixth factor can be predicted by the remaining five. All of the VIF scores exceeded 5.0 except for the presentation ability factor, which
was 4.87. This result confirms those found in the discriminant validity section and suggests that the six constructs do not appear to be sufficiently distinct from one another.

To move forward and assess the true factor structure of the 34 items, we decided to subject the 34 items to exploratory factor analysis. Because the items were highly correlated, direct oblimin rotation was used. Two factors emerged from the analysis. The first factor, which we label learning environment, is essentially a combination of the first three Brightman et al. (1993) dimensions: presentation ability, organization and content, and grading and assignments. The second factor, which we label student involvement, is a combination of the three remaining Brightman et al. (1993) constructs: intellectual and scholarly, student interaction, and student motivation. This result is similar to the two-factor solution that Goldstein and Benassi (2006) found. Only one item, the second on the original student motivation construct, cross-loaded (i.e., had a loading of over .4 on both factors) and thus was dropped from further analysis. We confirmed the reliability, unidimensionality, and discriminant validity of these two new factors. The results are presented in Table 2.

**Reliability and Validity—SERVQUAL Scale**

The steps for developing the SERVQUAL scale mimicked those for the Brightman scale. The results of the analysis are presented in Table 3. These results suggest that the five dimensions of SERVQUAL are reliable, unidimensional, and divergent. However, before proceeding to multiple regression, we also tested VIF scores for the

### Table 2: Development of two new factors.

<table>
<thead>
<tr>
<th>Factor (No. of Items)</th>
<th>Cronbach Alpha</th>
<th>First Eigenvalue</th>
<th>Second Eigenvalue</th>
<th>Minimum Factor Loading</th>
<th>Percent Var. Explained</th>
<th>Alpha—AVISC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning environment (24)</td>
<td>.97</td>
<td>14.3</td>
<td>.9</td>
<td>.581</td>
<td>59.3</td>
<td>.39</td>
</tr>
<tr>
<td>Organization &amp; clarity (9)</td>
<td>.94</td>
<td>5.4</td>
<td>.9</td>
<td>.503</td>
<td>59.6</td>
<td>.36</td>
</tr>
</tbody>
</table>

### Table 3: Factor development of the SERVQUAL scale.

<table>
<thead>
<tr>
<th>Factor (No. of Items)</th>
<th>Cronbach Alpha</th>
<th>First Eigenvalue</th>
<th>Second Eigenvalue</th>
<th>Minimum Factor Loading</th>
<th>Percent Var. Explained</th>
<th>Alpha—AVISC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance (4)</td>
<td>.89</td>
<td>2.9</td>
<td>.6</td>
<td>.776</td>
<td>74.0</td>
<td>.42</td>
</tr>
<tr>
<td>Empathy (4)</td>
<td>.94</td>
<td>4.0</td>
<td>.4</td>
<td>.835</td>
<td>79.8</td>
<td>.48</td>
</tr>
<tr>
<td>Responsiveness (3)</td>
<td>.92</td>
<td>2.6</td>
<td>.2</td>
<td>.932</td>
<td>86.5</td>
<td>.46</td>
</tr>
<tr>
<td>Tangibles (4)</td>
<td>.82</td>
<td>2.8</td>
<td>.6</td>
<td>.769</td>
<td>69.8</td>
<td>.53</td>
</tr>
<tr>
<td>Reliability (3)</td>
<td>.92</td>
<td>2.6</td>
<td>.3</td>
<td>.903</td>
<td>85.9</td>
<td>.43</td>
</tr>
</tbody>
</table>
five dimensions. The highest VIF score was for reliability at 3.24. All other scores were less than 2.0. These results confirm the appropriateness of using multiple regression.

RESULTS AND DISCUSSION

Hypothesis 1—Predicting Student Satisfaction with Course

Multiple regression analysis was used to determine if either of the two scales could predict a student’s overall satisfaction with the course, the dependent variable. The first model tested the two revised constructs derived from the study by Brightman et al. (1993). The standardized betas for learning environment and student involvement were $.478 (p < .01) and $.284 (p < .05), respectively. The overall model was significant at $p < .001$ and had an adjusted $R^2$-squared value of .489. This finding confirms Hypothesis 1a—student evaluations, measured using the revisions to the Brightman scale, are positively associated with student satisfaction with the course. Furthermore, these results confirm those of earlier studies that have also shown that organization and clarity and presentation ability (two of the three dimensions in our learning environment construct) have the strongest effect on student satisfaction (Feldman, 1989; Cabrera et al., 2001).

Next, the same hypothesis was tested using the five SERVQUAL dimensions as independent variables. The results of the multiple regression are shown below in Table 4. At the $p < .05$ level, three of the five SERVQUAL dimensions were positively related to student satisfaction with the course: empathy, reliability, and assurance. The other two dimensions, responsiveness, and tangibles, were not significant. The adjusted $R^2$-squared value for this model, .472, is very similar to that of the model which used the Brightman measures. This finding confirms Hypothesis 1b—student evaluations, measured using SERVQUAL, are positively associated with student satisfaction with the course.

Hypothesis 2—Predicting Student Satisfaction with Instructor

Similar to the methods used to test Hypothesis 1, we employed multiple regression to test Hypothesis 2, using the revised Brightman scales as the independent

Table 4: Multiple regression results, dependent variable: student satisfaction with course.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Std. Beta</th>
<th>t Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>.197</td>
<td>1.99</td>
<td>.048</td>
</tr>
<tr>
<td>Empathy</td>
<td>.606</td>
<td>3.23</td>
<td>.002</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>-.090</td>
<td>.567</td>
<td>.455</td>
</tr>
<tr>
<td>Tangibles</td>
<td>.016</td>
<td>.210</td>
<td>.834</td>
</tr>
<tr>
<td>Reliability</td>
<td>.289</td>
<td>2.36</td>
<td>.019</td>
</tr>
</tbody>
</table>

$F$ value = 19.829, $p < .001$.

$R^2$-squared = .472.
Table 5: Multiple regression results, dependent variable: student satisfaction with instructor.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Std. Beta</th>
<th>t Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>.268</td>
<td>2.19</td>
<td>.031</td>
</tr>
<tr>
<td>Empathy</td>
<td>.470</td>
<td>3.48</td>
<td>.001</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>-.060</td>
<td>-1.10</td>
<td>.272</td>
</tr>
<tr>
<td>Tangibles</td>
<td>-.060</td>
<td>-1.11</td>
<td>.271</td>
</tr>
<tr>
<td>Reliability</td>
<td>.336</td>
<td>3.13</td>
<td>.002</td>
</tr>
</tbody>
</table>

F value = 54.795, p < .001.
R-squared = .717.

variables first. The standardized betas for learning environment and student involvement were .642 (p < .001) and .252 (p < .05), respectively. The model had an F statistic of 94.3, which is significant at the p < .001 level. The adjusted R-squared value for this model was .727. Note that this value is much higher than the R-squared found in the tests to Hypothesis 1. This finding seems reasonable, as there are many contextual variables outside of the control of the instructor that may affect a student’s satisfaction with the course in general, such as the difficulty of subject, the time slot of course, the dynamism between students, the classroom, and so on. Similar to the findings in Hypothesis 1, the results of this model demonstrated that satisfaction with instructor is more heavily influenced by learning environment than student interaction, although both are significant. These results confirm Hypothesis 2a—student evaluations, measured using the Brightman scale, are positively associated with how satisfied those students are with their instructors.

Comparable results were found when the five SERVQUAL dimensions were used to measure students’ perceptions. The results are summarized in Table 5. This model found the same three dimensions significant, listed here in decreasing order of magnitude: empathy, reliability, and assurance. The other two variables remained nonsignificant. The overall model was significant at p < .001, confirming Hypothesis 2b—student evaluations, measured using SERVQUAL, are positively associated with how satisfied those students are with their instructors. The adjusted R-squared for this model, .717, is near that found in Hypothesis 2a, illustrating that the two student evaluation scales explain roughly the same amount of variance in a student’s satisfaction with the instructor.

Hypothesis 3—Predicting Student Perceptions of Learning

When the two revised Brightman et al. (1993) constructs were regressed against student perceptions of learning, only learning environment was a significant predictor. Its standardized beta of .560 is significant at the p < .001 level. Student involvement, standardized beta of .140, t statistic of 1.00, is not significant at the p < .10 level. The overall model F value was 47.2, which is significant at the p < .001 level. The adjusted R-squared for this model was .504. The lower R-squared value for student learning when compared to satisfaction with instructor is expected, as prior research has revealed that a vast array of important variables beyond the
control of the instructor affect student learning: student motivation, ability, personality, and so on. (Syler et al., 2006). These regression results serve to confirm Hypothesis 3a—student evaluations, as measured through the revised Brightman constructs, are positively associated with student perceptions of learning.

The results of the SERVQUAL model are depicted in Table 6. In this model, only two dimensions were significant predictors of student learning at the \( p < .05 \) level: empathy and assurance. The three other variables were all nonsignificant. The overall model is significant at \( p < .001 \) and had an adjusted \( R^2 \)-squared value of .508; again, this closely resembles the adjusted \( R^2 \)-squared in the model from Hypothesis 3a. These results confirm Hypothesis 3b—student evaluations, as measured through SERVQUAL, are positively associated with student perceptions of learning.

### Hypothesis 4—Correlation Between Expected Grade and Student Evaluations

As discussed earlier, one major concern with the use of student evaluations is the issue of grade expectation bias; in other words, are students who are expecting to receive high grades biased in their evaluations (Eiszler, 2002)? A good student evaluation instrument should not exhibit this bias. To test for this correlation, two preliminary steps must be taken. First, for each of the two major student evaluation scales one overall service quality score was devised. We achieved this in three different ways: the average score of all items (33 for Brightman, 19 for SERVQUAL), average score of regression factor scores (three scores for Brightman, five scores for SERVQUAL), and average score of exact factor scores (three scores for Brightman, five scores for SERVQUAL). All three methods produced nearly identical results: not only did significance results remain unaltered, parameter estimates were all extremely close, and as a result, only the details for the average score method are given. Second, as Centra (2003) pointed out, when testing for a correlation between student evaluations and expected grade, it is necessary to control for student learning. This was done by using partial correlations. The partial correlation, which controls for student learning, between expected grade and student evaluation, as measured through the Brightman scale, is \(- .049\), which is not significant at the \( p < .10 \) level. Similar results are produced when using the SERVQUAL scale to calculate a total student evaluation score; the partial correlation is \(- .083 \) (\( p > .10 \)). These findings suggest that once student learning is controlled for, expected grade

### Table 6: Multiple regression results, dependent variable: amount learned.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Std. Beta</th>
<th>( t ) Statistic</th>
<th>( p ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>.313</td>
<td>2.03</td>
<td>.043</td>
</tr>
<tr>
<td>Empathy</td>
<td>.406</td>
<td>2.26</td>
<td>.025</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>(- .180)</td>
<td>(- 1.04)</td>
<td>.300</td>
</tr>
<tr>
<td>Tangibles</td>
<td>.005</td>
<td>.07</td>
<td>.946</td>
</tr>
<tr>
<td>Reliability</td>
<td>.193</td>
<td>1.36</td>
<td>.178</td>
</tr>
</tbody>
</table>

\( F \) value = 23.839, \( p < .001 \).  
\( R^2 \)-squared = .508.
has no significant relationship with student evaluations—in other words, students expecting good grades do not necessarily rate instructors more highly than those expecting lower grades.

CONCLUSION AND FUTURE RESEARCH

This research is the first to apply the SERVQUAL scale to measure student perceptions of service quality in a classroom setting. Although the scale itself is well established, the application of it to the classroom is innovative in nature. The findings suggest that the SERVQUAL scale is reliable and exhibits both convergent and divergent validity. In fact, in terms of scale development, SERVQUAL performed better than a traditional student evaluation scale, the Brightman scale. In addition, the SERVQUAL scale has been shown to display predictive validity, because a significant positive relationship exists between individual dimensions of SERVQUAL and two measures of student satisfaction as well as student learning. Indeed, with far fewer questions (44% less), the SERVQUAL scale explains roughly the same amount of variance in these student outcome measures as does the Brightman scale. This parsimony may help prevent respondent fatigue and in the long run lead to a more reliable assessment instrument.

In predicting the various student outcome measures, the behavioral dimensions of SERVQUAL were the most effective. The dimension of empathy consistently had the strongest impact on the dependent variables. The items for this dimension, items largely ignored in the Brightman scale, suggest that students are looking for customer-centric qualities in instructors; in other words, an instructor who understands the individual needs of each student and is able to give personalized attention. A second SERVQUAL dimension was significant in all three models as well—assurance. This dimension shows that the students must feel confident in both instructors’ knowledge of their fields and their impartiality in assessment. The reliability dimension played a significant role in two of the three regression models, indicating that both consistency and dependability are traits that affect student satisfaction. Neither tangibles nor responsiveness had a significant effect in any of the three regression models, suggesting that these two traits are not as important as the three other SERVQUAL dimensions in shaping satisfaction and learning. Instructors can use these findings, that is, the regression rankings, as a resource allocation tool when devising teaching strategies and trying to improve the quality of their classroom experiences.

Collectively these results demonstrate that a customer-centric service quality scale such as SERVQUAL can be applied in a classroom setting. Although universities have been widely considered to be as close as possible to what is considered a pure service, and students are progressively seeing themselves more and more as consumers of said service, there has been little effort to blend existing service management literature into current academic research streams (Oldfield & Baron, 2000). Because the effects of improving university service radiate and multiply throughout the service value chain, this study’s findings are wide ranging indeed. As such, innovative methods must be devised to capture the student’s voice as active participant and customer in a service delivery encounter. As this study has demonstrated, one such method is using scales such as SERVQUAL. This
customer-centric approach can help instructors improve their service delivery, thus increasing service quality for many of the stakeholders in the education model: students obtain higher-quality classroom experiences, instructors receive information for professional development, the university gains a better reputation, future employers will get better trained graduates, and so on. Rather than focusing solely on many of the structural elements of the classroom experience, as many student assessment scales do, the SERVQUAL scale focuses on the behavioral aspects of the classroom. So in addition to trying to improve structural components such as the syllabus, outlines, handouts, exams, and so on, instructors can use the SERVQUAL assessment scale to understand what behavioral traits they need to improve and the latent construct names themselves are very intuitive, powerful, and easy to understand, particularly for business instructors who are familiar with service quality terminology. These abstract level constructs, terms such as empathy and assurance, certainly give the instructor a different perspective of the needs of their customers (students) than do questions asking about very specific components like pace of instructor’s speech, clarity of handouts, and so on. For example, results of student surveys might show instructors that they need to improve their overall responsiveness to their customers (students). The instructor could do so by extending office hours, checking e-mail, and/or phone messages more often and making use of chat tools in software programs such as WebCT and Blackboard. In general, instructors need to think of themselves as service providers in a common business sense. They can learn more about improving their service delivery when they incorporate these tacit customer-centric ideas and terminology into their assessment paradigms than by focusing on traditional scales that heavily emphasize very specific structural components of service delivery.

At this point, it is worth mentioning the limitations and extensions of this research. The respondents to this survey were primarily business majors enrolled in Operations Management classes. The findings should not be generalized until confirmed in a variety of settings. This study could be replicated in other departments within business schools as well as across entirely different disciplines. Likewise, this study focused on courses that were composed primarily of juniors and seniors, future research can extend these results to lower-level courses as well as graduate courses. Not only should the validity and reliability of the SERVQUAL instrument be validated, but so should the strength of the relationships between the five individual dimensions and the dependent variables. For example, this study found that empathy and assurance were the most significant predictors of satisfaction and learning. Perhaps in different environments, the other dimensions may be discovered to be more important. For example, in computer or science lab courses, tangibles might be a significant predictor of satisfaction and/or learning.

A second limitation is the use of self-assessment to gauge student learning and expected grade. Because this survey was anonymous, it was impossible to link student responses with exam scores or grades earned. However, immediately before the survey was distributed, students were given detailed sheets of all their scores to date with a summary at the end indicating what their grades would be if class ended that day. Self-reported scores have been used and shown to be reliable in previous research (e.g., Cabrera et al., 2001). Another potential methodological limitation is the possibility of common method variance exerting undue influence on the data set. To assess common method, we used Harman’s Single-Factor test.
Because of the limited sample size, we only tested the 19 SERVQUAL items as well as the 4 student outcome measures. Common method variance is assumed to exist if a single factor emerges from the unrotated factor solution and/or the first factor explains the majority of the variance in the items (Malhotra et al., 2006; Podsakoff & Organ, 1986). When the 23 items are subjected to exploratory factor analysis, 6 factors emerge, the 5 SERVQUAL dimensions as well as 1 factor of outcome measures. The first factor accounts for only 33% of the variance in the data. Taken together these findings suggest that the data do not exhibit extreme common method variance.

As discussed earlier, in addition to testing the generalizability of these findings, another potential area for future research is to investigate the uniqueness of the two scales. For example, is the SERVQUAL scale explaining the identical variance in the dependent variables (satisfaction, learning) that the Brightman scale is? Using the terminology introduced by Goldstein and Benassi (2006), SERVQUAL focuses primarily on the “process” portion of service delivery; can the SERVQUAL items be combined with the “structure” items included in Brightman’s scale to create an even more comprehensive educational service quality instrument? Future researchers can explore, possibly using hierarchical regression, the overlap between the two scales—are elements within the scales unique or complementary? In addition, some of the structural items in Brightman’s scale, such as organization and clarity, may actually be antecedents to the behavioral items in SERVQUAL. For example, clear organization might allow instructors to perform more reliably or become more responsive to student needs.

A close parallel to the idea presented above is to determine if the two scales can be parsimoniously combined in some manner to form one global scale maximizing the amount of variance explained in student outcome measures. Perhaps what one scale lacks, the other includes. In a similar vein, future researchers can determine whether a more limited set of SERVQUAL items could be used to measure the service quality in higher education. This study used 19 items; future research can explore whether either a subset of those 19 items or newly created items can explain as much variance in student outcomes variables as the scales used in this research.

Finally, this study used a general measure of student learning as a dependent variable. Learning has been shown to be multidimensional in nature. Future research can investigate the relationship between SERVQUAL and specific dimensions of learning such as professional competencies, group interpersonal skills, problem-solving skills, design skills, and so on (Cabrera et al., 2001). This type of research would aid in showing whether any additional questions should be added to the SERVQUAL scale in order to capture instructor behavior that can influence these specific types of learning dimensions. Likewise, another dependent variable that could be introduced could be aimed at teaching gains. Does the SERVQUAL scale, and the resulting analysis the instructor would receive from using it, lead to instructors improving their classroom performance?

REFERENCES


**APPENDIX A—QUESTIONS USED TO MEASURE SERVQUAL**

**Empathy**

1. The instructor is genuinely concerned about the students.
2. The instructor understands the individual needs of students.
3. The instructor has the student’s best long-term interests in mind.
4. The instructor encourages and motivates students to do their best.

**Assurance**

1. The instructor is knowledgeable in his/her field.
2. The instructor is fair and impartial in grading.
3. The instructor answers all questions thoroughly.
4. I am confident the instructor has an expert understanding of the material.
Responsiveness
1. The instructor quickly and efficiently responds to student needs.
2. The instructor is willing to go out of his or her way to help students.
3. The instructor always welcomes student questions and comments.

Reliability
1. The instructor consistently provides good lectures.
2. The instructor is dependable.
3. The instructor reliably corrects information when needed.

Tangibles
1. The classroom is modern and updated.
2. The physical environment of the classroom aids learning.
3. The classroom is equipped with all the necessary equipment to aid learning.
4. The classroom is kept clean and free of distractions.

Michael Stodnick is an assistant professor of operations and supply chain management in the Department of Management at the University of North Texas. He earned his PhD in management sciences from The Ohio State University. After receiving his BA, he worked as a materials manager for an axle company in Elkhart, Indiana. His research interests include service operations management, continuous improvement programs, and materials management. He is a member of and has presented his research at the Decision Sciences Institute, Institute for Operations Research and the Management Sciences, and Production and Operations Management Society.

Pamela Rogers is a doctoral candidate at the University of North Texas majoring in operations and supply chain management. She earned a BBA in Economics and Marketing and an MS in Computer Education and Cognitive Systems. Her business experience includes manufacturing, inventory control, computer-based training, and Web-site development as well as military training. Her research interests include manufacturing and service flexibility, supply chain flexibility, and operations strategy formulation. She has presented her research at the annual meetings of the Decision Sciences Institute, Southwest Decision Sciences Institute, and Institute for Operations Research and the Management Sciences.