clear that when Person One is allowed no commodity X and no commodity Y (i.e., the utility of Person One is zero), Person Two will reach a higher indifference curve $\mu'$ under the production possibility curve $WW'$ than the indifference curve $\mu'$ reached under the production possibility curve $ZZ'$. Therefore, the situation utility possibility curve derived from the altered production possibility curve $WW'$ will lie entirely outside that derived from the initial production possibility curve $ZZ'$.

An Approach to Standardization of Faculty Evaluations: An Empirical Study

DONALD SALYARDS and KEITH R. LEITNER

Introduction

The student unrest that pervaded the nation's campuses in the late 1960's triggered strong pressures for academics to inquire into their teaching effectiveness. As a result of this movement, student evaluations of teachers in many colleges and universities have become a very important factor in making decisions concerning faculty retention, promotion, and tenure. Such evaluations have to be analyzed with considerable care and deliberation since a number of complexities arise in regard to their interpretation.

Leniency in grading is one of the most powerful methods used by some teachers to raise their evaluation scores. With formalized student evaluations of faculty now standard practice, teachers may be tempted to use easy grading as a way to buy high student evaluations. Since leniency in grading may distort faculty evaluations, the purpose of this paper is to develop a way to standardize raw faculty evaluation scores for the grading system.

Although a number of studies have attempted to investigate the relationship between grade inflation and faculty evaluations, no consensus has been reached. While some studies have concluded that a strong positive relationship exists, others have suggested the relationship is weak at best. A major reason for this confusion is that the problem has not been conceptualized adequately and thus there is little theoretical guidance for the development of empirical studies.

Theoretical Model

To assess accurately the effects of grading leniency on faculty evaluations, consider the following analysis. Assume that the higher the utility the student acquires from attending a class under one professor, the higher the rating of the instructor and course. Furthermore, if students are rational, more leisure and higher grades are preferred to less leisure and lower grades. Suppose a student's utility function is:

$U = U(G, I, SL)$

where $G$ is the final course grade obtained, $L$ is the amount of time used for other courses and leisure activities, and $SL$ is a subjective measure of student learning from this course. All three variables affect the course utility directly. For the purpose of this study, the variable $SL$ is assumed to be of negligible importance. This assumption is based on the fact that when evaluations are administered in introductory or general studies courses, students appear not to know how much they have learned. A priori reasoning is supported by a recent study conducted on students at the introductory level at the University of Western Ontario.

The student maximizes his utility function subject to the transformation constraint:

$$L + \frac{I_0}{2} G = I_0$$

where $I_0$ is the maximum time available to the student over the whole semester after rest and self-care and $I_0/2$ is the constant opportunity "time price" of a unit of grade. Initially the student is faced with a transformation constraint depicted by line $L_0$ in Figure 1.

The student has various options ranging from an "A" in the course and $I_0$ leisure to an "A" in the course and no leisure. Given the indifference curve $I_0$, the utility-maximizing distribution of time between work and leisure is a grade of "C" and $I_0$ of leisure. No other allocation of time is preferred to this one, given the information and assumptions of the model. The effects of leniency in grading can now be analyzed. For example, if the teacher eases up, by making examinations easier, omitting more rigorous material, and/or lowering the grading scale, the transformation constraint shifts to the position $L_N$. However, only the solid section of $L_N$ is relevant. The student may work harder and end up with a grade higher than otherwise. Such a choice is shown by point $A_0$, where the student earns an "A" in the course by spending more time working on it. Also, the student may choose the same amount of leisure but a grade of "B" rather than "C" as before. This option is denoted by point $B_0$ on line $L_N$. Or, the student may prefer more leisure and the same grade as before. This combination is illustrated by point $C_0$. Finally, the student may choose an increasing amount of leisure time at the expense of a lower grade, as at $D_0$. In each of the four possible outcomes, the faculty evaluation score has risen since the student's relative utility has improved. The student is on a higher indifference curve, and chooses point $A_0$, $B_0$, $C_0$, or $D_0$, depending upon the utility function. The implication is that a change in the grading standard will lead to higher faculty evaluation scores but that such scores may have a positive, negative, or zero correlation with student grades. Thus, the relationship between these scores and student grades cannot be predicted on an a priori basis. This lack of predictability accounts for the mixed results of the empirical studies which regress student ratings of the instructor on their course grades.

Since a faculty member's evaluation score is related directly to that person's grading system, some way to standardize raw evaluation scores to assess teaching effectiveness more accurately must be developed. Based upon the above analysis, such a method must capture the effect of both student grades and time spent on the course to account completely for any alteration in the grading system.

Suppose that, in the absence of any inflation behavior, $G^*$ is the final course grade that a student would have earned by spending a total of $T$ hours in various activities relating to the course during the semester. Then $L = T$ would be the time used for all other purposes, including studying for other courses and leisure activities. The measure of faculty evaluation by the student then would have been $FE^*$. The observed or post-inflation counterparts of these variables are denoted by $G$, $T$, $I_0$, and $FE$ respectively. When the faculty member gives concessions, the student may choose a higher grade with a higher time-input, a higher grade and the initial time-input, the same grade and a smaller time-input, or a lower grade with a proportionally smaller time-input. Suppose the outcome is an inflation of the student's $G^*/T^*$ ratio by a factor $\alpha (\alpha > 1)$. In response, the student inflates the faculty evaluation by the factor $\alpha$. Therefore, $E = \alpha (G^*/T^*)$.

**FE** = $\alpha FE^*$

From (1) and (2),

$$E = \frac{FE}{G/T} = \frac{FE}{G^*/T^*}$$

where $E$ is the corrected faculty evaluation by a student according to this method. The variable $G/T$ can thus be used as an index for standardizing the raw faculty evaluation $(FE)^*$ in obtaining an estimate of $(FE^*/G^*/T^*)$.

The scale of this standardized evaluation is, however, quite different from the one for raw faculty evaluations $FE$ or $FE^*$. Suppose $FE$ and $FE^*$ vary over the range 0 to 4, $G$ varies from 0 to 4, and $T$ varies say from 0 to 100 hours. Then $FE/(G/T)$ would vary from 0 to 400. The individual observations of $FE/(G/T)$ obtained from one or all sections of a particular course could be arranged in a frequency distribution. The population mean, variance, and other characteristics of this distribution could then be estimated. The difficulty with this procedure is that complete anonymity of the students is hard to maintain. Each student's $FE$ is identified by his or her $G/T$. If this method is used, some way to maintain anonymity would have to be devised.

An alternative procedure is to calculate the ratio between the mean $FE$ and mean $G/T$:

$$E = \frac{FE}{G/T}$$

where $E$ is the corrected faculty evaluation by the whole group, $FE$ is the mean raw faculty evaluation, and $(G/T)$ is the mean $G/T$ for the group taking a course. It may be noted that the mean of $E$, will not generally be the same as $E$. This makes no difference as long as the approach selected is employed consistently. One disadvantage of this method is that a lot of information is lost in averaging before calculating the ratio. However, one advantage is the anonymity it assures.

Suppose now that the subjective learning variable $SL$ is a significant determinant directly affecting the student's utility from the course. As grade leniency tends to raise the $G/T$ ratio, it tends to lower $SL$, which causes an offsetting reduction in course utility. The ratio $G/T$ is inflating by the factor $\alpha$, the student's utility, and hence $FE$ is inflating by factor $\beta$, where $\beta < \alpha$. Since $\beta < 1$ would mean $FE < FE^*$, it would imply that the adverse effect on a student's utility through $SL$ is either stronger than or exactly offsetting...
the favorable effect of a rise in $G/T$. Both of these cases can be ruled out because it is assumed that the instructor will not allow the subjective learning to deteriorate to the point of getting the same or lower evaluation than the true one. Therefore, it is reasonable to expect that the true faculty evaluation distortion factor is $b < 1$, which is somewhat smaller than $a$. Thus the standardized faculty evaluation $E_2$ as defined in (3) will become:

$$E_2 = \frac{FE - \beta}{\alpha (G/T^*)} \quad (5)$$

where the factor $\beta/\alpha < 1$, if $SL$ is a significant variable. No such general statement can be made for the standardized evaluation $E_1$ as defined in (4).

Whether $SL$ is or is not a significant variable will be tested by including $SL$ as one of the regressors during the empirical analysis for the study.

**Data Collection and Measurement of Variables**

Data for the study were collected from 167 students in eight Principles of Economics classes in three collegiate institutions during the Spring semester of 1977. Careful internal controls were used to ensure that data pertaining to $T$, $FE$, and $G$ were accurate as possible.

The collection of time-input data ($T$) required special effort. Time-input by a student included the number of hours actually spent on all activities relating to a course during the entire semester. Such activities include class attendance, library work visits to tutors and faculty, study at home or during classes at school. For accurate record keeping, a special time sheet showing days of the week and a list of different activities relative to course needs was prepared. Every week the faculty member distributed these sheets to the students and asked for their cooperation in recording daily time-input for each week. It was recorded to the nearest fifteen minutes.

It was stated clearly on the sheet that the data were being collected for research purposes only and would not affect the grade either way. However, student suspicion would be raised if an instructor knew the time spent by a student before a final course grade was assigned. Thus, to avoid false answers on time sheets, each student was assigned a special number for identification on the "name" portion of the time sheet. The number was assigned by a neutral person either a professor other than the instructor, or a student in the class. This person presented the instructor with a list matching student names to the special student numbers only after the final course grades had been submitted to the registrar. At the end of the semester, weekly time sheets for each student were added to yield the total number of hours spent by each student during the semester.

Earned course grades ($G$) of students were available from faculty records. They were classified on the usual 0 - 4.00 scale, with (+) and (-) values used to make the scale more precise. Data pertaining to the raw faculty evaluation ($FE$) and subjective learning ($SL$) were obtained from a 38 question, standard faculty evaluation form used in all eight classes. The raw faculty evaluation ($FE$) was quantified by adding the responses of students to 32 evaluation form questions. The subjective learning ($SL$) variable consisted of student responses to the remaining six questions on the standard faculty evaluation form. These pertained to students' self perceptions of learning in the course, and do not necessarily reflect actual student learning. All evaluation forms were given to students during the last week of classes, with students also indicating expected course grades at the time of evaluation.

**The Empirical Model**

The study uses a multiple regression model to examine the relationship between the faculty evaluation score ($FE$), achieved course grades ($G$), time-input ($T$) and subjective learning ($SL$). The importance of each independent variable is determined, ceteris paribus, by the values of the partial correlation coefficients. Standard statistical assumptions apply to the regression model, including linearity in the parameters, an additive relationship between the independent variables, a constant variance, and a normally distributed error term with an expected value of zero.

The model is specified as:

$$FE = a + b_1 G + b_2 T + b_3 SL + e$$

where $FE$ represents the raw faculty evaluation score, $G$ represents the earned course grade, $T$ represents the time-input, and $SL$ represents the subjective learning variable for the ith student. For equation (6) a positive relationship is hypothesized between $FE$ and the $G/T$ ratio, and an insignificant relationship is hypothesized between $FE$ and the $SL$ variable.

Some studies have suggested that the faculty evaluation score ($FE$) might be better related, not to the actual course grade received by the student, but to the student's expected grade at the time of the evaluation. Therefore, the above model was also tested substituting $G$ for $G_e$, where $G_e$ is the expected grade at the time of the evaluation as opposed to the actual course grade the student receives.

**Results and Conclusions**

Table 1 illustrates the regression results for the study. The first eight equations come from individual classes where actual grade per unit of time ($G/T$) was used as an independent variable. Equation 9 represents the results of

<p>| Table 1: Multiple Regression Coefficients for Subjective Learning and Grade/Time Factors |
|-----------------------------------------------|----------|----------|----------|----------|</p>
<table>
<thead>
<tr>
<th>Equation</th>
<th>Class Description</th>
<th>$R^2$</th>
<th>$R^2$</th>
<th>$R^2$</th>
<th>Subjective Learning</th>
<th>Grade Per Unit of Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Winona State Univ.</td>
<td>36</td>
<td>.51</td>
<td>17.2%</td>
<td>2.51</td>
<td>4.16</td>
</tr>
<tr>
<td>2</td>
<td>College of St. Teresa</td>
<td>26</td>
<td>.35</td>
<td>6.2%</td>
<td>3.23</td>
<td>1.03</td>
</tr>
<tr>
<td>3</td>
<td>Univ. of Wisconsin/Eau Claire #1</td>
<td>31</td>
<td>.71</td>
<td>22.5%</td>
<td>3.59</td>
<td>1.24</td>
</tr>
<tr>
<td>4</td>
<td>Univ. of Wisconsin #2</td>
<td>12</td>
<td>.70</td>
<td>16.7%</td>
<td>3.81</td>
<td>1.81</td>
</tr>
<tr>
<td>5</td>
<td>Univ. of Wisconsin #3</td>
<td>13</td>
<td>.66</td>
<td>9.7%</td>
<td>2.48</td>
<td>5.88</td>
</tr>
<tr>
<td>6</td>
<td>Univ. of Wisconsin #4</td>
<td>29</td>
<td>.66</td>
<td>4.5%</td>
<td>1.19</td>
<td>2.15</td>
</tr>
<tr>
<td>7</td>
<td>Univ. of Wisconsin #5</td>
<td>11</td>
<td>.31</td>
<td>1.8%</td>
<td>2.28</td>
<td>2.96</td>
</tr>
<tr>
<td>8</td>
<td>Univ. of Wisconsin #6</td>
<td>19</td>
<td>.74</td>
<td>22.5%</td>
<td>2.87</td>
<td>3.05</td>
</tr>
<tr>
<td>9</td>
<td>Aggregate ($G/T$)</td>
<td>167</td>
<td>.59</td>
<td>117.3%</td>
<td>3.02</td>
<td>3.55</td>
</tr>
<tr>
<td>10</td>
<td>Aggregate ($G_e/T$)</td>
<td>167</td>
<td>.58</td>
<td>113.4%</td>
<td>5.01</td>
<td>1.65</td>
</tr>
</tbody>
</table>

*The $R^2$ values appear beneath each coefficient and an asterisk denotes significance at the 5 per cent level.*
all eight classes combined, using G/T as an independent variable. Equation 10 aggregates all eight classes, but expected grade per unit of time (G/T) was used as an independent variable in place of the actual grade per unit of time (G/T). Equations for individual classes using expected grade per unit of time (G/T) were not included in the study because their results were not unlike the first eight equations in Table I.

"F" values were significant for nine of the ten equations in Table I, indicating that a significant amount of the total variance was explained by the model. However, the explanatory statistical significance of the subjective learning (SL) variable offered a surprising refutation of the original hypothesis that SL would be insignificant. On the other hand, the G/T variable, which was hypothesized to be positively related to faculty evaluation scores, was not significant in eight of the equations, and was just marginally so in the remaining two. The empirical evidence leads to two conclusions:

(1) The hypothesis that students in introductory courses are seeking a high grade per unit of time spent in the course, and that they rate instructors primarily on this criterion is false, and
(2) Students rate instructors according to their perception of actual learning during the course. Subjective learning seems to be a highly significant variable affecting students' perception of instructor performance.

The major emphasis of this study was to determine whether faculty evaluation scores can be deflated by the G/T ratio to provide a more accurate assessment of faculty performance. Apparently, deflating faculty evaluation scores by this ratio would serve no purpose because such a ratio does not have a significant positive relationship to faculty evaluation scores. The G/T ratio is statistically significant in equation 9, but is not significant in seven of the eight previous equations. Implementation of any deflation tool requires that it be done on a class-by-class basis; thus, even if G/T is statistically significant for the aggregation of all classes, it is not worthwhile to implement such a procedure if it does not apply to individual faculty class sections.

Rejection of the original hypotheses suggests that faculty evaluation scores need not be deflated by a G/T ratio to make them more indicative of actual teaching performance. Indeed, since the subjective learning variable is highly significant, the question might well be, "Does anything need to be done with raw faculty evaluation scores?" It should be remembered that the SL variable in this study is not a measure of actual learning (as in the Super study), but is a measure of students' perception of learning in the classroom.

This study has contributed to the literature because time has now been considered empirically as it determines (along with the grade) the faculty evaluation score. Apparently, such a G/T effect is not significant. While discovering what is not significant is not as exciting as finding something unusually novel and significant, it does illuminate the corridors of future research. The magnitude of the F* values for the above equations shows that factors other than subjective learning might also affect students' perception of instructor performance. Clearly, more research needs to be done to determine what other factors motivate a student to rate his or her instructor.

I. An Overview of Gibson's Approach

Gibson's theoretical approach is summed up in the following quotation:

"The economy is initially at point E0, where money stock is M0, and real (nominal) interest rates are r0. If income does not rise immediately following an increase in the money stock to M0, the economy will more immediately rise to point P, along the curve L1, drawn for a given level of income and expected rate of price change. This fall of interest rates from r0 to r1 is the liquidity effect. As income increases, the L curve will shift rightward until it reaches the position L0, which intersects M0 at E0, where interest rates are again at r0. We may call the movement from P to E0 the income effect, and it obviously just balances the liquidity effect. If now the expected rate of price change increases, L0 will shift to the left, lowering real interest rates to r0. Nominal rates will exceed real rates by the expected rate of price increase."

The Dynamics of Interest Rate Adjustment in Response to Monetary and Fiscal Changes

SULEMAN A. MOOSA

Univ. of Illinois, Chicago Circle