A STUDENT'S MOST SEVERE CRITIC

by Ken D. Duft*

*Assistant Professor, Department of Agricultural Economics, Washington State University, Pullman, Washington, with special statistical assistance from my colleague, Lee Blakeslee.

What factors contribute most towards student dissatisfaction with our existing letter or numerical grade system for judging exam performance? Is it that the grade or score ultimately becomes the student's whole raison d'etre? Is it that grades or scores place the student in neat little pigeon holes from which it becomes difficult to extricate himself? Is it that the process of grading or scoring an exam, itself, has little relevance to the learning process? Or, is it that student dissatisfaction is merely a reflection of student exam score expectations exceeding by an appreciable degree actual scores received? I postulate that most of my colleagues would probably argue that dissatisfaction actually results from all these factors and others even more subtle. In general, I would agree, except that in my view, the last factor noted above should be excluded.

Contentions

Since the beginning of my professional career, I have held a personal conviction that, despite some student claims to the contrary, once examined, student score expectations do not differ greatly from the scores actually received. More specifically, I believe that if facilitated, a student's self-appraisal of his examination answers will result in a score not unlike that actually levied by the instructor. In addition, I have maintained that if a difference did actually exist between an instructor's judgement and a student's own (referred to as the I-S variation), then the student's own self-made critique would probably be the most severe of the two. **Hence, if given an opportunity, a student is likely to become his own most severe critic.**

Two hypotheses have been made. First, I contend that the end result of our current grading system would change very little if students (rather than the instructor) were allowed to judge their own performance. Second, I contend that in such cases where a significant I-S variation is noticeable, the student's own appraisal would result in the lower of the two grades received. With each passing year, my desire to test the validity of my two related contentions has increased. Finally, this semester I decided to take the necessary steps to conduct such a test.

Methodology

A five-part mid-term examination (constituting 25 percent of the final course grade) was prepared as a standard part of my teaching responsibilities over a senior-level, 3-unit course in "agribusiness management." Each part of the exam was valued at 20 points and designed as a mini-case problem requiring a discussion-type answer. There existed no exact answer to any of the five questions, but instead, each part asked the student to propose and justify one or alternative solutions to common management problems. Hence, instructor evaluation of student performance would, unavoidably, be a subjective process. Nonetheless, I reviewed and scored each examination, part by part, being careful to make all notations and record all scores on a separate piece of paper. All exams were returned to the students unmarked. The students were informed that my evaluation process was complete, but that individual marks had not yet been entered in the grade book. I then explained that for the entire class period, we (students and instructor) would review the test questions, discussing each in detail, and finally reach an agreement as to what the answers to each part should approximate. Reaching such agreement was to result from a meaningful compromise between what I viewed as being a "good" answer and the students' personal views of same. Hence, both instructor and student views and biases became major (but equal) elements in the compromise answer. Once the agreed-upon answer became evident, each student would be asked to judge (and score from 0 to 20) his own performance on each exam part. However, before the scoring procedure began, students were informed that the actual score entered in the grade book would be either the instructor's score, or the student's own score, **whichever was the higher**. In so doing, the possibility of my placing an undeserved burden on the students (because of my passion to experiment) had, therefore, been eliminated. If, in turn, a bias was to result from this declaration, it would obviously be in support of a refutation of my second contention. Student and instructor scores for each of five parts and for the exam total were then assembled, summarized and statistically tested to accept or reject the following hypotheses: 1) the I-S variation for each of the five exam parts and for the exam's total score is statistically insignificant, and 2) for those exam parts where an I-S variation does exist, that variation will be positive and statistically significant.

Results

To test the validity of my first hypothesis the following statistical procedure was followed:

A null hypothesis which states that the I-S variation would be statistically insignificant for each exam part and the exam, as a whole, was subjected to a "t-test" evalution. The null hypothesis (and my first contention) would be accepted if the calculated t proved less than the table t value at the 5% level of significance. The statistical results of this evaluation are shown in the Appendix.

In general, however, Table 1 shows my first contention to be less than wholly confirmed. The null hypothesis is accepted for exam parts 2, 4, and 5, suggesting that for these exam parts, the I-S variation would approximate zero. However, in exam parts 1 and 3, and for the whole exam, student self-appraisal would result in scores significantly different from those of the instructor.

TABLE 1
Summary of Instructor and Student Scores With Statistical Interpretations

Statistic	Exam Parts 1	2	3	4	5	Whole Exam
Table $t_{.05}$	1.78	1.78	1.78	1.78	1.78	1.78
Cal. t	2.19	1.49	3.59	-.49	-.98	3.84
Accept Hyp.		√		√	√	
Reject Hyp.	√		√			√

Hence from this statistical evaluation, the I-S variation has not been shown to be insignificant in all cases. To accept or reject my second contention that the I-S variation is positive, an attempt was made to regress S on I. Then linear regression equation used was: $S_{ij} = a + bI_{ij}$

If the results of this regression show that the regression coefficient (b) is both significant (rejecting a hypothesis that b = o) and positive, and that the regression yields a high correlation coefficient (r, where $r = \frac{\Sigma IS}{\sqrt{(\Sigma \chi^2)(\Sigma I^2)}}$, then my second contention shall be accepted. The statistical results of this test are also shown in the Appendix.

In this case, the t-test evaluation supports my second contention in all exam parts as well as the whole exam (see Table 2). We must reject the hypothesis that b=0. Moreover, as shown in the Appendix, all regression coefficients (b) are arithmetically positive and all correlation coefficients are high (no lower than .704) for any exam part or for the exam as a whole.

TABLE 2
Summary of Regression Analysis of Student On Instructor Scores

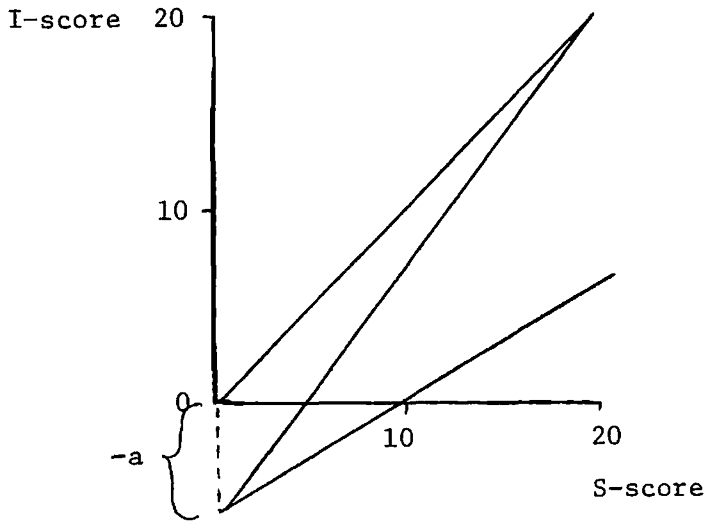
Statistic	Exam Parts 1	2	3	4	5	Whole Exam
Table $t_{.05}$	2.179	2.179	2.179	2.179	2.179	2.179
Cal. t	3.287	3.996	4.068	4.543	16.246	4.243
Accept Hyp.						
Reject Hyp.	√	√	√	√	√	√

From the above two statistical tests, one may conclude that in many cases students will tend to grade in agreement with their instructors. In those cases where student and instructor grades are in significant disagreement, the students will grade themselves more severely than will their instructor.

A Side Issue

The statistical results shown in the tables above and in the Appendix illustrate an important side issue worthy of some brief note. The reader will note that for most exam parts the I-S relationship produces a negative regression intercept ($a<o$) and a slope (regression coefficient, b) which varies above and below the value of one. If this I-S relationship were plotted (see Figure 1), a perfect 1:1 correlation would be depicted by a 45 degree line (labeled A). If a $b>1$ is derived from our evaluation, the I-S relationship can be depicted by line B, which indicates that the student performing well (by his own judgment) on a particular exam part will more closely approximate the instructor's score than will the student performing poorly on the same exam part. Stated a little differently, the student performing poorly on a particular question, is more likely to render a score (his own) which results in a larger I-S variation. Conversely, if $0<1$ is derived from our evaluation, our interpretation is reversed, as illustrated by line C. In fact, the data show this latter case to be true, i.e. $0<b<1$ on those exam parts where the students performed poorly. Hence it can be said that the student performing poorly (by his own judgment) on a particular exam part will more closely approximate the instructor's score than will the student performing well on the same exam part.

Figure 1
I-S Score Alternative Relationships



Limitations

The results above are in answer to a personal inquiry. The methodology evolved as a compromise between practicality, logic and intuitive appeal. The experiment, itself, was not applied to a representative sample, but to a particular class, using a specific exam, and conducted at a single institution. No attempt is made to suggest that the test results are sufficient to describe the character of all students in all classes at all institutions. Inductive generalizations, therefore, cannot be extended. Moreover, the experimentation process did involve a risk. That is, students were given the opportunity to artificially inflate their test score. In my opinion, in this case, all students conducted a conscienteous self-appraisal, recording that score deserved as opposed to that desired.

Value

As already noted, one value of the experiment, itself, is that it satisfied a personal curiosity. As such, it proved to be a personally satisfying activity. However, as a truly scientific inquiry, it proved to be little more than a simple statistical exercise. If true values were to evolve from the process, they had to be associated with benefits accrued to the students' ability to learn and/or the instructor's ability to teach. In this case, such values were noticeable. Indeed, the process by which the instructor and his students reached a negotiated answer to each exam part proved to be a valuable learning experience for the students. The students were given a "piece of the action" in that the final answer to each part contained both the instructor's choice of critical elements and those deemed equally important by the student. As each student compared his own individual answer with the negotiated one, a more vivid acknowledgement of his own shortcomings resulted. As such, the student also achieved experience in self-evaluation. Insofar as self knowledge is a teaching-learning goal, this procedure permitted the student to learn himself and face squarely the responsibilities associated with rendering a judgment.

From the instructor's point of view, the negotiation process enabled him to reiterate and defend relevant elements of those lectures which preceded the exam. Finally, the experiment's results proved personally satisfying as they suggested that once the goals of the exam questions were established (correct answers agreed upon), students illustrated not only that they are good evaluators, but that they will accept this responsibility in an objective manner.

STATISTICAL APPENDIX

First Null Hypothesis:

$H_o: \mu\chi_{ij} = 0$, where $\chi_{ij} = I_{ij} - S_{ij}$ and,
I_{ij} = instructor grade on the i^{th} exam part for the j^{th} student,
S_{ij} = the j^{th} student's own grade on the i^{th} exam part
μ = hypothesized mean
$\bar{\chi}$ = calculated mean
$i = 1, \ldots, 6$ for exam parts 1-5 and exam total 6
$j = 1, \ldots, 13$ for students participating

calculated $t = \dfrac{\bar{\chi} - \mu\chi_i}{\frac{S}{\sqrt{n}}}$, where $S = \sqrt{\dfrac{\Sigma\chi^2 - \frac{(\Sigma\chi)^n}{n}}{n-1}}$

n = 13

Statistic	Exam Parts 1	2	3	4	5	Whole Exam
Aver. Inst. Score	17.38	14.38	11.23	7.77	1.00	51.77
Aver. Student Score	15.08	13.38	8.31	8.62	1.62	47.00
I-S Variation	2.30	1.00	2.92	-1.15	- .62	4.77
$\bar{\chi}$	2.31	1.00	3.15	- .54	- .62	4.77
S	3.79	2.42	3.17	3.95	2.26	4.48

Second Null Hypothesis:

$H_o: b=0$, where $S_{ij} = a \pm bI_{ij}$
S_{ij}, I_{ij} = as before
Correlation coefficient = $r = \dfrac{\Sigma IS}{\sqrt{(\Sigma\chi^2)(\Sigma\chi^2)}}$

Statistic	Exam Parts 1	2	3	4	5	Whole Exam
a	-23.20	-.35	-.57	1.61	- .25	1.21
b	2.20	.96	.79	.90	1.87	.89
r	.70	.77	.76	.81	.98	.86