

---

# DETERMINANTS OF STUDENT FEEDBACK SCORES

---

By

*Willie Tan*  
*Department of Building*

*If you are teaching a lecture class, are older, has students with lower expected grades, teach in a relatively "low score" Department, and are teaching a large class, the odds appear to be stacked against you...*

## **Introduction**

Student feedback (SF) teaching scores are often used as a mode of assessing teachers at institutions of higher learning as well as for other purposes such as rewards and motivation (e.g. best teacher awards) and feedback to improve the quality of teaching. However, debates on SF scores tend to generate more heat than light, and one of the reasons is that every teacher seems to have his or her favorite theory. While we conceptualize in our individual silos and mostly from our own experiences, there is little hard evidence to test these theories so that appropriate actions may be taken.

The purpose of this paper is to present evidence on the determinants of SF scores. It is hypothesized that SF scores depend on:

- structural variables such as type of activity (lecture, seminar or sectional teaching), class size, module level, Department and Faculty in which the module is offered, and whether it is a General Education Module (GEM);
- teacher characteristics such as gender and age. The latter is a proxy variable for teaching experience. Although the number of years a teacher has taught a module is probably a better proxy variable, data are difficult to obtain. Besides, a teacher may join NUS much later in his life for various reasons; and
- student characteristics such as students' expected grade.

Obviously, one can think of many other variables such as quality of teaching materials, class composition, time of activity (day or night), self-selection by students to take a particular module, and classroom environment. The above variables were used primarily because data were readily available.

## Developing the model

Having decided on the variables to use, the next step was to design an appropriate estimating model. Clearly, naïve analysis using one or two variables will not suffice without proper statistical control of the compounding effects of other variables. For this reason, a multiple regression model was used. This technique correlates the dependent variable (SF scores) with the independent variables simultaneously. It is worth highlighting the well-known distinction between correlation and causation where, in addition to association between variables, a causal mechanism (i.e., reason) needs to be provided to demonstrate causality. However, if any two variables are not even correlated, then inferring causality based on regularity is out of question. There are, of course, many instances where causality does not depend on statistical regularity, such as the causes of World War II.

As a preliminary step, simple graphical plots revealed that the relationships between each independent variable and SF scores could be nonlinear. Consequently, a nonlinear model was used, and the details are provided in the Appendix for the benefit of readers who are more statistically inclined.

## What have we found?

Based on 1983 data points (modules) from Semester 1 of AY2004/5 offered at the National University of Singapore, the results are as follows:

a) The correlation between the percentage of nominations for teaching awards and SF scores was 0.61, which is not high. This is probably because SF scores reflect the majority view and this need not correlate with teaching excellence.

b) The following variables have *no significant impact* on SF scores at the university level:

- whether the module is a *GEM*;
- *module level* from Year 1 to Year 6 (PhD level); and
- *gender* of the teacher.

It is, of course, possible that these variables may have some effect at faculty or departmental levels. For instance, if class composition is heavily tilted towards a particular gender, then the gender of the teacher may matter.

c) The following variables have an impact on SF scores:

- whether a module is a *lecture or sectional teaching*, which are different modes of delivery. Lectures scored lower than sectional teaching;
- *students' expect grade*, which has a positive impact on SF scores. Students who expect higher grades tend to give higher SF scores.
- *Department or Faculty* that offers the module, which is not unexpected given different knowledge areas;

- *age* of teacher. Older teachers tend to have lower scores. Perhaps younger teachers can relate to students better, although teachers who are inexperienced are unlikely to fare as well. The effect of experience on SF scores is likely to be quadratic, that is, experience counts up to a certain number of years and then it becomes less important whether one has 20 or 25 years of teaching experience; and
- *class size* – for those who teach large classes, SF scores tend to fall nonlinearly with class size. Table 1 shows the rate of decline of SF scores with class size controlling for the effect of other variables. The curve is normalized at SF = 5.0 for a module with one student. The fall in mean SF scores with increasing class size is quite rapid for small class sizes. Thus, a teacher teaching a class of 300 can be expected to score, on average, about  $0.48 - 0.26 = 0.22$  lower than teaching a class of 20. For example, a score of 4.22 for a class of 20 is equivalent to a score of 4.00 for a class of 300.

Class size	Mean SF score	Difference
1	5.00	-
20	4.74	0.26
30	4.71	0.29
50	4.67	0.33
100	4.61	0.39
200	4.55	0.45
300	4.52	0.48

Table 1. Fall in SF scores with increasing class size.

## Conclusion

Student feedback scores are affected by the type of activity (lecture or sectional teaching), class size, students' expected grade, age of teacher, and the department or faculty in which the module is offered. In particular, class size has a major nonlinear effect on SF scores. Simple rules of thumb may be derived using Table 1.

Several caveats are in order. First, this paper merely presents some evidence on the determinants of SF scores at the aggregate (university-wide) level using a single set of data (i.e. one university) at one point in time. Relationships that hold at the university level may not hold at the faculty level. The selection of variables is arguably subjective and largely guided by data constraints. There is no data on many qualitative variables that could affect SF scores, and it is not surprising that the R-square as a rough measure of model fit is only 0.18, which is relatively low (although  $t$  values are significant).

Second, interpretation of the results is largely left to the reader to reexamine his or her favorite theories in the light of the evidence. One of the aims of this paper is to generate debates on the issues and it is clearly not possible to discuss in depth all the possible reasons for the findings.

Third, the implications of the results are a separate but important matter. After all, SF scores are only one indicator of teaching quality and an imperfect one. Many views have been expressed on this point such as the inability of some students to assess a teacher properly, failure to distinguish good teaching and good mentoring, mixing the “business” of teaching with friendship, penalizing teachers who are more demanding, and rewarding teachers who teach at an elementary level. There is also a “three-parts theory” where teachers with extreme SF scores (too high or too low) could be too unpopular on one hand and too popular on the other hand for the wrong reasons.

In summary, there is evidence to show that certain structural variables, teacher characteristics, and student characteristics affect SF scores. However, many of the qualitative factors that affect SF scores remain elusive. It is unwise to read too much into what is ultimately a simple model that gives us a feel of the possible determinants of SF scores. There is no need to split hair or engage in meaningless rankings over what is well known as at best a rough indicator of teaching quality. Needless to say, inappropriate uses of any performance measure affect morale and therefore performance.

## Acknowledgments

I would like to thank Prof KP Mohanan for his considerable assistance in developing the model and Mr Quek Jin Nan for his useful suggestions and in providing the data.

## Appendix

This appendix is intended for readers who wish to examine the statistical basis of the model. The model is

$$(1) \quad SF = C e^{\alpha LECT} e^{\beta SEMINAR} n^{\phi} AGE^{\gamma} e^{\mu GENDER} e^{\pi GEM} ML^{\psi} e^{\eta DEPT} e^{\kappa FACULTY} EXGRADE^{\lambda} \varepsilon$$

where  $C$  is a constant,  $e$  is the base of natural logarithm,  $\alpha, \beta, \dots, \lambda$  are parameters, and  $\varepsilon$  is the error term. Taking logs on both sides, we have

$$(2) \quad \ln(SF) = \ln(C) + \alpha LECT + \beta SEMINAR + \phi \ln(n) + \gamma \ln(AGE) + \dots + \ln(\varepsilon).$$

In addition, the percentage of nominations for best teacher award was also correlated with SF scores.

The variables are:

$SF$  – student feedback score for module;

$LECT$  – dummy variable; 1 = Lecture, 0 otherwise;

$SEMINAR$  – dummy variable; 1 = Seminar, 0 otherwise;

*Sectional Teaching* – coded as 0 for LECT and 0 for SEMINAR.

*n* – class size

*AGE* – age of teacher

*GENDER* – dummy variable; 1 = Male; 0 otherwise

*ML* – module level, integer (1-6);

*DEPT* – integer

*FACULTY* – integer

*EXGRADE* – mean students' expected grade

The estimated regression model is

$$(3) \quad \ln(SF) = 1.244 - 0.03014LECT - 0.0176\ln(n) - 0.0402\ln(AGE) - 0.00173DEPT \\ \quad \quad \quad \quad \quad \quad \quad (-5.41) \quad \quad \quad \quad (-9.59) \quad \quad \quad \quad (-3.61) \quad \quad \quad \quad (-4.63) \\ \quad \quad \quad \quad \quad \quad \quad + 0.00774FACULTY + 0.28865\ln(EXGRADE) \quad R^2 = 0.18 \\ \quad \quad \quad \quad \quad \quad \quad (3.36) \quad \quad \quad \quad (11.62)$$

The brackets below the estimated parameters are t-values that are significant at the 0.05 level of significance. Analysis of residual plots indicated no substantial departures from normality.

The effect of class size is examined further by partially differentiating Equation (1) so that

$$(4) \quad \frac{\partial E[SF]}{\partial n} = \frac{\phi E[SF]}{n}$$

where E[.] denotes expectation. A discrete approximation is

$$(5) \quad \phi = \frac{\% \Delta SF}{\% \Delta n} .$$

Thus,  $\phi$  represents elasticity, the percentage change in SF score divided by the percentage change in class size.

For practical purposes, we may write the estimated equation as

$$(6) \quad E[SF] = An^{-0.0176}$$

where *A* represents the effects of all other variables. If we normalize (6) so that  $SF = 5.0$  for  $n = 1$ , we obtain Table 1 in the main paper.

Apart from the low R-square, the model uses integer codes for Departments and Faculties to save on degrees of freedom but this requires the assumption of a constant difference in SF scores between Departments and Faculties, which may not be realistic.