Failing Students Optimally

Arthur J. Caplan
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/eri

Recommended Citation
https://digitalcommons.usu.edu/eri/291
Economic Research Institute Study Paper

ERI #2004-17

FAILING STUDENTS OPTIMALLY

by

ARTHUR J. CAPLAN

Department of Economics
Utah State University
3530 Old Main Hill
Logan, UT 84322-3530

November 2004
FAILING STUDENTS OPTIMALLY

Arthur J. Caplan, Assistant Professor

Department of Economics
Utah State University
3530 Old Main Hill
Logan, UT 84322-3530

The analyses and views reported in this paper are those of the author(s). They are not necessarily endorsed by the Department of Economics or by Utah State University.

Utah State University is committed to the policy that all persons shall have equal access to its programs and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

Information on other titles in this series may be obtained from: Department of Economics, Utah State University, 3530 Old Main Hill, Logan, UT 84322-3530.

Copyright © 2004 by Arthur J. Caplan. All rights reserved. Readers may make verbatim copies of this document for noncommercial purposes by any means, provided that this copyright notice appears on all such copies.
FAILING STUDENTS OPTIMALY

Arthur J. Caplan

ABSTRACT

This paper estimates the deadweight loss associated with not failing an optimal number of students. We find that this loss ranges between $3,200 and $4,600 per student over the student's four-year undergraduate career. One possible ramification of adopting a more stringent student-failure policy (to recoup the deadweight loss) is investigated. We find that an institution can promote greater effort on the part of both students and faculty by encouraging faculty to fail more students.

JEL Classification: D21, D61, I20

Key words: optimal fail rate, student-failure policy, deadweight loss
Failing Students Optimally

Acknowledgements
Thanks to Glenn Davis, Utah State University (USU) Registrar, and Caryn Beck-Dudley, Dean of USU College of Business, for granting access to the student transcript data used in this study, and Randy Shelton for compiling the data. Prithvi Jutur was instrumental in writing the Excel macros that further compiled the data into the dataset used for this study. John Gilbert wrote the macro which enabled the completion of the bootstrapping exercises. I also greatly benefited from discussions with colleagues David Dickinson, David Aadland, and John Gilbert; discussions that ultimately provoked me to write this paper.

1. Introduction
The on-going debate over grade inflation is focused on whether instructors have been too generous in awarding their students A and B grades over the past 25 years and thereby inflating cumulative grade point averages. On one side of the debate, Rojstaczer (2004), Johnson (2003), Rosovsky and Hartley (2002) and Levine (1998) provide evidence of significant grade inflation occurring at virtually every public and private institution nationwide. On the other hand, Adelman, et al. (2003) argue that only minor changes in grade distributions have occurred. Kohn (2004) claims that the rising tide of grades does not in itself prove that grade inflation exists—it must be proven that higher grades are undeserved. This debate is important from an economic standpoint because, as Ostrovsky and Schwarz (2003) point out, years of grade inflation make grades permanently less informative, leading to potential labor-market distortions. Grade
inflation also presents a prisoner’s dilemma for institutions that might otherwise consider implementing a grade-deflation policy. Their students might be disadvantaged in the labor market relative to students from other institutions.

The debate, important as it is, has left unanswered an enticing question. Rather than asking whether we are awarding students too many A’s and B’s, the better question might be, are we awarding enough F’s? The case in favor of more F’s is compelling. Failing more students may increase an institution’s profitability. Profits would increase, for instance, if the tuition revenue obtained from the next student who retakes a failed course exceeds the sum of (i) the expected tuition lost if the student instead decides to drop out of the institution rather than retake the course, and (ii) any opportunity costs associated with the loss of excess institutional capacity (e.g., potential bottlenecks in course enrollments or new-student admissions). Thus, a get-tough policy may make economic sense regardless of whether the institution is experiencing grade inflation. To the extent that it desires to maximize profit, the institution would be rational to seek an optimal student failure rate. Furthermore, common knowledge about the institution's

---

1 Failing more students might also exacerbate the grade inflation problem for institutions that not only replace an F grade with the new grade once the student retakes the course, but also recalculates the student’s cumulative grade point average without the earlier F grade included.

2 This firm-orientated perspective of course abstracts from the more market-orientated prisoners-dilemma problem alluded to earlier.

3 In cases where its grading standard—reflected here as optimally choosing the number of courses students will typical fail—is positively correlated with its prestige, a prestige-maximizing institution with current grading standards (i.e., number of student failures) beneath its optimal level will necessarily move toward the prestige optimum by choosing instead to maximize profits as long as current grading standards are similarly beneath the profit optimum. Given the nationwide trends in state funding of public institutions, profit-maximizing behavior at prestige-minded institutions is also more likely to become the rule over time rather than the exception (Lohmann, 2004; Riggs, 2004; Ehrenberg, 2000; Steinberg, 2000; and Lenth, 1993).
new “get-tough” student-failure policy may induce students and professors to devote
more effort to their studies and course preparations, respectively.\textsuperscript{4,5}

This paper presents a simple theoretical framework within which an optimal
decision rule is derived for the number of failed courses per student (Section 2) and
provides empirical estimates of (i) the optimal number of failed courses per student and
(ii) the corresponding deadweight loss (Section 3). We find that the optimal number of
failed courses per student is approximately six. Given that the typical student currently
fails slightly less than one course during his four-year career, this translates into a
deadweight loss of between $3,200 and $4,600 per student. Section 4 presents the
heuristics of how a get-tough student-failure policy might induce both students and
professors to devote more effort toward the learning experience and thereby move from a
low-effort/low-effort to a high-effort/high-effort state of the world. Section 5 concludes.

2. The Theory of Failing Students Optimally

Consider a profit-maximizing academic institution that chooses the number of courses
typical students of class-standing $i$ majoring in a discipline in each of $j$ colleges will fail
during the period of their degree programs, $F_{ijt}$.\textsuperscript{6} The institution’s problem may be
written as,

\textsuperscript{4} Imagine the laughs an institution’s public relations office could have while designing a
marketing campaign for the new get-tough policy. A few slogans that quickly (and perhaps imprudently)
come to mind are “F ‘em,” “F You,” and “Just say no to grade inflation.”

\textsuperscript{5} Professors may have greater incentive to invest more effort in their courses if they are no longer
encouraged to withhold F’s for students who they might otherwise not have passed. Failing grades might
be withheld in order to (1) avoid having to deal with the students for a second time (Knight Ridder Tribune
Business News, 2003), (2) maintain future enrollments (Dobbs, 2004), or (3) maintain good student
evaluations (Arnold, 2004; Johnson, 2003; Cohen, 2003; O’Dell, 2003; Walzer, 2003; Rosovsky and
Hartley, 2002; and Becker and Watts, 1999). Of course, the incentive to withhold failing grades is
dependent upon the commensurate policies adopted by the institution to alleviate these three concerns.

\textsuperscript{6} In- versus out-of-state students, domestic versus foreign students, and specific major are three
other possible student identifiers that might be important to administrators. However, it is unlikely that
subject to

\[ S_i = \overline{S} - \sum_{j=1}^{J} \alpha_j C_j \left( \sum_{i=1}^{I} F_{ijt} \right) \geq 0, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J \]  

where \( T_t^F \) is the per-course tuition rate (assumed constant across student types and colleges) in period \( t \); \( T_{it}^{CD} = \int_{t}^{t+T_i} T_t^D e^{-\rho t} \) is the discounted cumulative tuition revenue lost for student type \( i \) (assumed constant across colleges) from period \( t \) onward (\( T_t^D \) is tuition revenue lost during period \( t \) and \( T_i \) is the finite number of periods during which tuition revenue is lost); \( P_{ijt} \) is student \( i \)'s (in college \( j \)) probability of dropping out of the university (\( P_{ijt}^e > 0, P_{ijt}^c \geq 0 \) \( \forall i, j, \) and \( t \)); \( \rho > 0 \) is a discount rate; \( S_t \geq 0, \overline{S} > 0 \), and \( C_j \) are, respectively, current excess institutional capacity, initial excess institutional capacity, and lost excess capacity at college \( j \)'s level (\( C_j^e > 0, C_j^c \geq 0 \) \( \forall j \)); \( \phi > 0 \) is the per-unit value of excess institutional capacity (measured in implicit revenue); and \( 0 < \alpha_j < 1 \) is a college-level capacity weighting factor, \( \sum_{j=1}^{J} \alpha_j = 1.7 \).

The institution’s problem is therefore to choose (i) a vector of the number of failed courses and (ii) excess institutional capacity that maximize the difference between

\[ \text{Max} \int_{t=0}^{T_i} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} (T_t^F F_{ijt} - P_{ijt} (F_{ijt}) T_{it}^{CD}) + \phi S_i \right] e^{-\rho t} dt, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J \]

these types of identifiers would proffer permissible distinctions upon which to base a student-failure policy (e.g., it seems unrealistic to assume that an administration could specify distinct optimal failure rates for in-versus out-of-state students, domestic versus foreign students, or electrical-engineering versus communications majors as they might for freshmen versus seniors and majors in the college of engineering versus the college of humanities, arts, and the social sciences).

Note that \( \rho > 0 \) along with the fact that the profit function is continuous, nonnegative (by the assumption of nonnegative profits in the long-run), and bounded (by the second-partial condition on \( C_j \)) ensures that the integral in the objective function will converge (Chiang, 1992).
the discounted stream of aggregate revenue (based on the sum of additional tuition receipts from students choosing to retake failed courses and the implicit value of excess institutional capacity) and the discounted stream of expected costs (based on expected cumulative future tuition receipts foregone as students choose to dropout rather than retake additional failed courses), subject to an excess institutional-capacity constraint. The rate at which excess capacity diminishes over time is in turn governed by each college’s weighting factor (i.e., the rate at which the loss of a unit of excess capacity at the college’s level impacts the university’s overall excess institutional capacity) and the degree to which the number of failed courses incurred by the typical student impacts the respective colleges’ excess capacity levels.

For example, if new student enrollment in a critical number of colleges is particularly sensitive to the aggregate number of courses failed by students \( i = 1, \ldots, I \) due to potential classroom- or faculty-availability constraints (e.g., \( C_j \) is relatively large at the given \( F_{ij} \) for a critical number of colleges) and these colleges utilize a relatively large share of the institution’s overall capacity (i.e., \( \alpha_j \) for these colleges is also relatively large), then, all else equal, courses failed by typical students in those colleges will have a relatively large negative impact on the institution's overall excess capacity. As shown below, this negative effect on institutional excess capacity in turn increases the incremental cost associated with an additional failed course, which ultimately works to reduce the optimal number of courses failed by student \( i \) in college \( j \), \( \forall i \) and \( j \).

Assuming an interior solution for number of failed courses (i.e., \( F_{ij}^* > 0 \) \( \forall i \) and \( j \)), the necessary conditions for this maximization problem are (1) and (dropping the \( t \) subscript for convenience),
\[ T_i^F = P_{ij}^r T_{ij}^{CD} + \alpha_j \phi C_j', \quad i = 1, \ldots, I, \quad j = 1, \ldots, J \]

Conditions (2) are the marginal decision rules determining \( F_{ij}^* \), \( i = I, \ldots, I \), \( j = I, \ldots, J \).

In this case, \( T_i^F \) represents marginal revenue from an additional failed course by student \( i \) in college \( j \) and \( P_{ij}^r T_{ij}^{CD} + \alpha_j \phi C_j' \) represents the associated (expected) marginal cost. The first marginal cost term, \( P_{ij}^r T_{ij}^{CD} \), accounts for the cost associated with lost cumulative future tuition revenue due to an increased probability that the student will drop out. The second term represents the added cost associated with lost excess institutional capacity. It reflects the possibility that failing more students will create bottlenecks in course enrollments or administration that may in turn preclude the institution's ability to admit new students at or above the historical rate.

If the role of excess capacity in determining \( F_{ij}^* \) is determined (or, due to data limitations, is believed) to be minor, e.g. because of negligible values for \( \phi, \alpha_j \) or \( C_j' \) (for a critical number of colleges), then the necessary conditions for this problem collapse to

\[ T_i^F = P_{ij}^r T_{ij}^{CD}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J. \]  

(2')

Figure 1 depicts condition (2') for student \( i \) in college \( j \).

If this student is historically failing fewer than \( F^* \) courses during the period of his degree program, the institution would experience a gain in profit if the student were issued more failing grades. Figure 1 therefore illuminates the main empirical question. If \( F^* \) can be empirically estimated, it can be compared with the actual number of courses the typical student fails to determine whether the institution should promote a more rigorous student-failure policy.
3. The Empirics of Failing Students Optimally

In order to estimate \( F^* \), transcript data on 22,000 undergraduate and graduate students were obtained from Utah State University (USU) Registrar’s Office based on a sample frame of all students who had declared a major in the College of Business (COB) either prior to or during 1995-2004, regardless of whether the student ultimately graduated with that or another COB major, and who was either still enrolled during that period or had graduated or dropped out.\(^8\) Students who were still enrolled, who transferred out of COB to a different college at some point during their studies, or for whom we had missing data were subsequently dropped from the sample. This resulted in a sample size for the empirical analysis of approximately 13,700 students who had either graduated with a COB degree or who had dropped out of the institution.\(^9\) Restricting the sample in this way has enabled the joint estimation of simple probit and negative-binomial models to explain the relationship in the data between the probability of a student dropping out and the number of courses failed, i.e., \( P' \). This information, along with current tuition rates assessed by the institution, in turn allows us to estimate \( F^* \).

Table 1 includes the definitions and summary statistics for the variables used in the empirical analysis. The mean value for \( \text{DROP} \) indicates that 66 percent of our sample is defined as having dropped out rather than having graduated during the period 1995-2004. This statistic reflects the liberal definition of how dropping out is defined for this study. Any student who at some point during their COB degree program took a leave of absence for at least two consecutive semesters is classified as having dropped out. It is

\(^8\) The Registrar was initially apprehensive about releasing data for this type of study. With subsequent written approval from the Dean of the COB, the Registrar agreed to provide access only to COB’s transcript data. Because of data limitations for \( \phi \), \( \alpha_j \) and \( C_j \), the estimation of \( F^* \) for this study is based on (2') rather than (2).

\(^9\) As will be discussed below, our measure of “dropping out” for this study is a liberal one.
likely that several of these students either later returned to complete their studies in a major offered by another college or eventually plan to return to complete their degrees in the future. Unfortunately, we are unable to distinguish with our data this type of student from the one who drops out of the institution for good. As a result, if DROP is biasing our estimate of $F^*$ in any direction, it is biasing it downward. This is due to the fact that the more students who drop out for a given number of failed courses, the steeper is the marginal cost curve ($P'T_{CD}$) in Figure 1. For this reason, our estimate of $F^*$ should be considered a lower-bound on the true optimal number of failed courses.

The mean value for FAIL? (0.35) indicates that approximately one-third of the students fail at least one course during their degree program. Although the average student only fails approximately one course during their program (mean of #FAIL = 0.88), he nevertheless retakes approximately two non-failed courses (mean of RETAKE = 1.78). Lastly, note that the mean value of AGE for our data (30.56) is skewed slightly to the right due to the relatively large number of older students who typically return to the university to complete a business-related degree.

To determine $F^*$, a two-step, or limited-information maximum-likelihood (LIML) procedure is used to jointly estimate $P'$, i.e., $Pr(DROP = 1)$, and #FAIL (Greene, 2003). Joint estimation is required not because of any theoretical justification for the sequential nature of the DROP and #FAIL outcome, but rather because of the obvious potential for endogeneity that exists between these two variables from an econometric standpoint. In other words, although we know that students first experience the #FAIL outcome before deciding the DROP outcome—i.e., #FAIL is predetermined with respect to DROP—as the

---

10 Intercooled Stata 7.0 for Windows 95/98/NT was used to produce the empirical results reported in this section.
econometrician we are precluded from treating #FAIL as a pre-determined variable. From our perspective, or as an artifact of the data-gathering process, DROP and #FAIL occur simultaneously. Because of this, and the strong likelihood that the stochastic processes underlying DROP and #FAIL are jointly determined, joint estimation of the two variables is justified.

In the first step we estimate \( \Pr(\text{#FAIL}|X_1) \), where each student's #FAIL is drawn from a negative-binomial distribution to account for the count nature of the data with mean over-dispersion. The covariate vector \( X_1 \) includes a subset of the variables included in Table 1.

Table 2 presents our results for the estimation of \( \Pr(\text{#FAIL}|X_1) \). We note that, on average, older and male students are predicted to fail more courses than their respective counterparts, and that the more courses a student takes the greater the number of courses he is expected to fail. To the contrary, married students and those students with higher GPAs are predicted to fail fewer courses than their counterparts. From the summary statistics for this regression we note that the data are not distributed Poisson (the null hypothesis of no over-dispersion (\( \alpha = 0 \)) is rejected at the 1 percent level of significance).

From the estimation of \( \Pr(\text{#FAIL}|X_1) \) we obtain a vector of predicted values for #FAIL (henceforth denoted \( \hat{F} \)), which is used as a covariate in the second-step LIML estimation of the student's probability of dropping out. Here, we estimate \( \Pr(\text{DROP} = 1|X_2) \) assuming a standard normal (probit) distribution conditioned on covariate vector \( X_2 \) with robust standard errors (White, 1980). To correct the estimator of the covariance matrix for this second-step regression (due to \( \hat{F} \) being used in place of
#FAIL), we perform a series of 500 bootstrapped regressions for both steps one and two in order to recover a non-parametric estimate of the standard errors in step two. To perform this procedure, we take the predicted values from the step-one bootstrapped regressions and sequentially “feed” them into the step-two regressions. The standard deviations of the resulting distributions of the coefficients are then used as our estimates of the step-two standard errors.  

Table 3 reports the estimated marginal effects for \( Pr(DROP = 1 | X_2) \). We note that, on average, older students are more likely to drop out than younger students. To the contrary, men, married students, students that have been enrolled for more semesters, students with higher GPAs, seniors, and graduate students are less likely to drop out. In comparing the marginal effects of \( \hat{F} \) and \( \hat{F}^2 \) on \( Pr(DROP=1) \) we indeed obtain an upward-sloping marginal cost over a range of failed courses, as depicted in Figure 1. The goodness-of-fit measures, \( \Omega_1 \) and \( \Omega_0 \), indicate that the model correctly predicts 94% and 80%, respectively, of those students who dropped out and those who did not drop out. These measures count as a correct prediction any predicted value that is within a magnitude of 0.5 of its corresponding DROP value.

To determine \( F^* \), the marginal effects for \( X_2 \) are combined with 2004-2005 tuition data obtained from the USU Budget Office. Excluding the additional expenses associated with room and board, books and supplies, and other personal expenses, but including student-body fees, resident and non-resident (undergraduate) tuition costs per

---

11 See Efron and Tibshirani (1993) and Stine (1990) for further information regarding this bootstrap procedure.

12 The positive marginal effect on the interaction term (FAIL?)\( (GPA) \) indicates that the negative effect of GPA is less for those students who have failed at least one course.

13 Further information on how the marginal effects and tuition data are combined is available upon request from the author.
three-credit course are estimated to be $421 and $1,212, respectively. The university reports that approximately 80% of the incoming freshmen for 2004-2005 were residents. Therefore, a weighted average of the resident and non-resident tuition costs based on this percentage of resident students results in a per-course tuition cost (i.e., $T^F$) of approximately $580.

The USU Budget Office also reports annual tuition costs of approximately $3,330 and $9,700 per year (two semesters) for resident and non-resident undergraduate students, respectively. For the purposes of this study, we therefore assume that the (discounted weighted-average) cumulative tuition foregone to the university when seniors decide to drop out (presumably at the beginning of their senior year) equals approximately $4,464. For ease of estimation, we lump freshmen, sophomores, and juniors into one class (non-seniors) and assume that the decision to drop out for non-seniors is made with two years remaining to complete their degree program. Based on the yearly tuition costs cited above, the (discounted weighted-average) cumulative tuition foregone to the university when non-seniors decide to drop out equals approximately $8,796. Assuming that the percentages of freshman, sophomores, juniors, and seniors at any given time are each 25%, the average cumulative tuition foregone is $7,713.

Figure 2 presents the (constant) marginal revenue of an additional failed course ($T^F$) plotted at $580, the current average number of failed courses per student enrolled in

---

14 These figures are based on a typical student enrolling for four three-credit courses per semester. For example, the Budget Office's reported tuition cost per semester for a resident is $1,686, implying a $1,686 / 4 = $421.50 estimated cost per course. Similarly, for non-residents based on a reported per-semester tuition cost of $4,850.

15 The annual tuition costs for resident graduate students are only slightly higher than for undergraduate students. We therefore proceed with the assumption that undergraduate and graduate incur equal annual tuition costs and that if a graduate student decides to drop out of the university, s/he does so at the beginning of the second year of study. In this way, tuition foregone from graduate students is identical to that from seniors. We assume a 3% ($\rho = 0.03$) discount rate.
the COB plotted as a thicker vertical line at 0.88 (see Table 1), and, based on the empirical results contained in Table 3, the expected marginal costs associated with lost future cumulative tuition revenue from seniors \( (\text{PT}^{\text{CD}}(\text{Senior})) \) and the average student \( (\text{PT}^{\text{CD}}(\text{Average Student})) \). The expected marginal cost for the average student is calculated by applying the marginal-effect estimates contained in Table 3 to the corresponding mean values (from Table 1) of the variables included in \( X_2 \) (excluding \( \hat{F} \) and \( \hat{F}^2 \)). Note that optimal number of failed courses for the average student is approximately six, implying a per-student deadweight loss (over the course of the student's degree program) of approximately $3,200 (area A).

Similarly, the expected marginal cost for a senior is calculated by applying the marginal-effect estimates to the mean values of the variables in \( X_2 \), except for SENIOR and GRAD, both of which are set equal to one. This results in an optimal number of failed courses of approximately eight and a corresponding deadweight loss of approximately $4,600 (area A+B). These two sets of results imply that there is scope to fail the average senior an additional two courses during his senior year.\(^{16}\)

4. One Possible Ramification of Failing Students Optimally

A get-tough grading policy of the type investigated here may motivate both students and professors to devote more effort to their studies and course preparations, respectively. Therefore, the learning experience at a given institution could potentially move from a low-effort/low effort to a high effort/high effort state as depicted in Figure 3. As shown in the appendix, as long as the student's preferences are (positively) monotonic over some

\(^{16}\) For graduate students, this translates into failing the average student approximately two courses during his two-year degree program.
initial range of effort and the professor’s preferences are similarly monotonic in teaching effort, the possibility of such a change in the state of the world—motivated by a change in the institution’s grading policy—indeed exists.

We assume that the professor’s preferences are more complex than the student’s, reflecting the fact that promotion (and thus utility) is based on a weighted average of research and teaching performance (the latter of which is based at least partially on student teaching evaluations). To the extent that teaching performance is based on student evaluations, and student evaluations are in turn based on course grades, the institutional change in grading policy will likely need to be accompanied by a change in the weights given to student evaluations. For example, at the same time students are learning about the new grading policy, the institution might consider informing faculty that their teaching performance will now be based less on student evaluations and more on peer review of their teaching portfolios. By enervating the influence of student evaluations on the assessment of teaching performance, professors may have greater incentive to enhance the rigor of their courses (i.e., increase their effort levels) commensurate with the increased effort levels of (the now more fearful?) students.

Of course, the potential moral-hazard problems of overzealous faculty failing too many students and overwhelmed (at the possibility of failing more courses) students could unravel the high effort/high effort outcome. However, it would seem that such an unraveling effect would be motivated more by the former moral-hazard problem, and thus would be amenable to potentially inexpensive forms of monitoring.
5. Conclusions

This paper has provided answers to two simple questions. First, are we in the teaching profession failing an optimal number of students? Our results suggest that the answer is no. There is ample scope for the typical student to fail more courses over the course of his academic career, resulting in a recovery of between $3,200 and $4,600 in deadweight loss per student. Second, what are the possible ramifications associated with adopting a more stringent student-failure policy in order to achieve the optimal failure rate? We discuss one scenario where both the students and professors increase their effort levels and thus mutually enhance the learning experience.

Future research should focus on providing additional evidence in support or in refutation of the results presented here. With respect to the first question—Do we fail enough students?—applying the approach used in this paper to alternative datasets would test the robustness of the answer provided here. With respect to the second question—what are the possible ramifications?—empirical evidence as to (i) the effects of professor effort on student effort, and vice-versa, and (ii) the inhibiting effects on professors’ grading decisions of an institution’s heavy reliance on student evaluations would enable a test of the assumptions made in this paper. In the end, evidence may indeed mount in favor of a more stringent student failure policy; a policy that not only increases the institution’s profitability, but also motivates more rigor in the learning experience itself.
References


Arnold, Roger A. (2004) "Commentary; Way That Grades Are Set is a Mark Against Professors; Awarding Students A’s for C-Plus Work Robs the Best and Brightest." Los Angeles Times, page B13, April 22.


Appendix

Begin by considering the student’s problem,

$$\begin{align*}
\max_{\{e_s^i, x_s, l_s\}} & \quad u^s(e_s, x_s, l_s) \\
\text{subject to,} & \\
M_s &= p_s x_s = w_s L_s \quad \text{(A1)} \\
T_s &= L_s + l_s + e_s \quad \text{(A2)} \\
e_s &= \Omega_s \left[ \gamma(\theta) e_p^i + \left[ 1 - \gamma(\theta) \right] e_s^i \right] \quad \text{(A3)}
\end{align*}$$

where $e_s^i$ is the component of overall student effort, $e_s$, that “emanates from within” the student and $e_p^i$ is the professor's teaching effort level. Thus, in (A3) $e_s$ is a weighted average of inherent-student and professor-induced effort levels, where the weights, $\gamma$, are functions of the student’s innate distribution parameter, $\theta$, and a university leverage factor $\Omega_s > 0$ ($= 1$ in the case of neutral leverage). Constraint (A1) is the income-budget constraint, where $M_s$ is student income level, $x_s$ is a composite good with associated price $p_s$, and $L_s$ is labor supplied with associated wage rate $w_s$. Constraint (A2) is a time constraint, where total time available is represented by $T_s$ and leisure time by $l_s$. Finally, utility $u^s$ is increasing and quasi-concave in both $x_s$ and $l_s$, but may be parabolic in $e_s$ (increasing up to some threshold $\bar{e}_s$ and decreasing beyond).

After some tedious algebra and assuming sufficient second-order conditions hold, it can be shown that although $\frac{\partial e_s^i}{\partial e_p^i} < 0$ (i.e., that students will substitute professor-induced effort for their own innate effort) their overall effort level, $e_s$, may nevertheless respond
positively to the professor’s teaching effort level. In other words,

\[
\frac{\partial e_s}{\partial e_p} = \Omega_s \gamma(\theta) + [1 - \gamma(\theta)] \frac{\partial e_s}{\partial e_p} > 0 , \text{ where } \Omega_s \gamma(\theta) > 0 \text{ is the direct effect of professor teaching effort on overall student effort and } [1 - \gamma(\theta)] \frac{\partial e_s}{\partial e_p} < 0 \text{ is the indirect effect.}
\]

Therefore, as long as the direct effect outweighs the indirect effect over some initial range of \(e_p\), the student’s effort reaction function will be positively sloped over that range.

One possible shape of the student's effort reaction function is depicted in Figure A1 below (as curve \(e_s\)).

**Figure A1.** Possible Student and Professor Effort Reaction Functions

---

17 To show this result, use the fact that if \[\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc < 0, \text{ then for any } t > 0, \]

\[\begin{vmatrix} -ta & c \\ -tb & d \end{vmatrix} = bc - ad > 0, \text{ which can be generalized to a square matrix of any dimension.}\]
Unfortunately, the solution to the professor’s problem is not nearly as clean. It can be written as,

$$\text{Max} \{ \{e'_p, e''_p, x_p, l'_p \} \} \ u^s(e'_p, e''_p, x_p, l'_p; e_s)$$

subject to,

$$M_p = p_p x_p = W_{p1} + \Pr(\text{Eval}) W_{p2} \quad (A4)$$

$$\text{Eval} = \Lambda'_p e'_p + \Lambda''_p e''_p + (1 - \Lambda'_p - \Lambda''_p) \overline{SE} \quad (A5)$$

$$\overline{SE} = S\big(f(e_s), g\big(f(e_s), e'_p)\big) \quad (A6)$$

$$T_p = l'_p + e'_p + e''_p \quad (A7)$$

where $e'_p$, $e''_p$, $x_p$, $l_p$, $M_p$, $p_p$, and $T_p$ are defined analogously to the student’s problem, with $e'_p$ representing the professor’s research effort level. In (A4), $W_{p1}$ is a base salary level, $W_{p2}$, is an increment to $W_{p1}$ based on the outcome of an institutional evaluation of the professor’s overall performance, Eval represents the outcome of the evaluation, and $Pr(\text{Eval})$ is the expected outcome of the evaluation, where $Pr$ is a subjective probability density function defined over Eval. In (A5), Eval is defined as a weighted average of teaching and research effort, as well as the average score from student evaluations, $\overline{SE}$, where $\Lambda'_p$ and $\Lambda''_p$ are the corresponding weighting factors. In (A6), $\overline{SE}$ is defined as a function of the distribution of student effort, $f(e_s)$, and the distribution of students’ grades, $g$, which is in turn based on the students’ and professor’s (teaching) effort levels. $S$ is assumed strictly increasing in $g$, but potentially parabolic in $f(e_s)$, while $g$ is assumed strictly increasing and concave in $f(e_s)$ and $e'_p$. Finally, (A7) is the professor’s time constraint.
Given the complexity of the professor's problem, the sign of $\frac{\partial e^1_p}{\partial e_s}$ is indeterminate.

Figure A1 therefore depicts one of any number of possible professor effort reaction functions, denoted $e^1_p$. As depicted, the low effort/low effort state of the world (point A) is unstable – indicating a divergent effort path in either direction. However, the high effort/high effort state at point B is steady without oscillation from either direction. In this situation, any tweak away from point A in the direction of point B will move the effort equilibrium steadily to B.
**Table 1.** Variable Definitions and Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENTER</td>
<td>Year student first enrolled at USU (21 = 2004, ..., 1 = 1983).</td>
<td>13.51 (2.49)</td>
</tr>
<tr>
<td>DROP</td>
<td>1 = dropped out, 0 = graduated</td>
<td>0.66 (0.47)</td>
</tr>
<tr>
<td>AGE</td>
<td>Age (in years) when first enrolled at USU.</td>
<td>30.56 (22.07)</td>
</tr>
<tr>
<td>GENDER</td>
<td>1 = male, 0 = female</td>
<td>0.59 (0.49)</td>
</tr>
<tr>
<td>MARITAL</td>
<td>1 = married, 0 = single</td>
<td>0.35 (0.48)</td>
</tr>
<tr>
<td>ETHNIC</td>
<td>1 = international, 2 = Asian, 3 = white, not of Hispanic origin, 4 = other, 5 = black, not of Hispanic origin, 6 = other Hispanic</td>
<td>2.86 (0.67)</td>
</tr>
<tr>
<td>#SEMS</td>
<td>Total number of semesters enrolled during degree program.</td>
<td>9.47 (6.10)</td>
</tr>
<tr>
<td>#COURSES</td>
<td>Total number of different courses taken during degree program.</td>
<td>31.00 (19.60)</td>
</tr>
<tr>
<td>STAND</td>
<td>1 = undergraduate student, 2 = graduate student, 3 = second bachelors degree</td>
<td>1.13 (0.37)</td>
</tr>
<tr>
<td>FAIL?</td>
<td>Did student fail at least one course during degree program?</td>
<td>0.35 (0.48)</td>
</tr>
<tr>
<td>#FAIL</td>
<td>Total number of courses failed during degree program.</td>
<td>0.88 (1.79)</td>
</tr>
<tr>
<td>GPA</td>
<td>Cumulative GPA at end of degree program (if FAIL? = 0)</td>
<td>3.04 (0.57)</td>
</tr>
<tr>
<td>RETAKE</td>
<td>Number of courses retaken that were not failed.</td>
<td>1.78 (3.12)</td>
</tr>
<tr>
<td>SENIOR</td>
<td>Class standing when either dropped out or graduated.</td>
<td>0.5 (0.5)</td>
</tr>
<tr>
<td>GRAD</td>
<td>Class standing when either dropped out or graduated.</td>
<td>0.14 (0.35)</td>
</tr>
</tbody>
</table>
Table 2. Regression Results for Pr(#FAIL|X_i).

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Coefficient Estimate (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>-18.82 (138.63)</td>
</tr>
<tr>
<td>STAND</td>
<td>-0.25** (0.05)</td>
</tr>
<tr>
<td>AGE</td>
<td>0.002*** (0.0008)</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.08*** (0.03)</td>
</tr>
<tr>
<td>MARITAL</td>
<td>-0.06** (0.03)</td>
</tr>
<tr>
<td>ETHNIC</td>
<td>-0.01 (0.02)</td>
</tr>
<tr>
<td>#COURSES</td>
<td>0.01*** (0.001)</td>
</tr>
<tr>
<td>FAIL?</td>
<td>19.62 (138.63)</td>
</tr>
<tr>
<td>GPA</td>
<td>-0.04* (0.02)</td>
</tr>
<tr>
<td>α</td>
<td>0.21*** (0.01)</td>
</tr>
<tr>
<td>( \chi^2(\alpha) )</td>
<td>632.80***</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-7549.26</td>
</tr>
<tr>
<td>( \chi^2(LR) )</td>
<td>13,045***</td>
</tr>
<tr>
<td>Pseudo R^2</td>
<td>0.46</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>11,862</td>
</tr>
</tbody>
</table>

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.
Table 3. Regression Results for $\Pr(DROP = 1|X_2)$.

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Marginal Effects (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAND</td>
<td>-0.15*** (0.015)</td>
</tr>
<tr>
<td>ENTER</td>
<td>-0.05*** (0.002)</td>
</tr>
<tr>
<td>AGE</td>
<td>0.02*** (0.0007)</td>
</tr>
<tr>
<td>GENDER</td>
<td>-0.03*** (0.007)</td>
</tr>
<tr>
<td>MARITAL</td>
<td>-0.06*** (0.008)</td>
</tr>
<tr>
<td>ETHNIC</td>
<td>-0.03*** (0.005)</td>
</tr>
<tr>
<td>#SEMS</td>
<td>-0.008*** (0.001)</td>
</tr>
<tr>
<td>GPA</td>
<td>-0.13*** (0.01)</td>
</tr>
<tr>
<td>(FAIL?)(GPA)</td>
<td>0.10*** (0.013)</td>
</tr>
<tr>
<td>SENIOR</td>
<td>-0.30*** (0.01)</td>
</tr>
<tr>
<td>GRAD</td>
<td>-0.69*** (0.03)</td>
</tr>
<tr>
<td>$\hat{F}$</td>
<td>-0.25* (0.14)</td>
</tr>
<tr>
<td>$\hat{F}^2$</td>
<td>0.05*** (0.01)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-3165.86</td>
</tr>
<tr>
<td>$\chi^2$(LR)</td>
<td>8186.67***</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.56</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>11,860</td>
</tr>
</tbody>
</table>

$\Omega_1 = \frac{\text{Predicted DROP } = 1}{\text{Observed DROP } = 1}$

$\Omega_0 = \frac{\text{Predicted DROP } = 0}{\text{Observed DROP } = 0}$

***Significant at the 1% level. *Significant at the 10% level.
Figure 1. The Optimal Number of Failed Courses (Theory).
Figure 2. The Optimal Number of Failed Courses (Empirically).

![Diagram showing the optimal number of failed courses for average and senior students. The diagram includes a line graph with the x-axis labeled "Number of Failed Courses" and the y-axis labeled "Dollars." The graph shows the optimal failure points for average and senior students.]

$P^T_{CD}(\text{Average Student})$ $P^T_{CD}(\text{Senior})$
Figure 3. The Student/Professor Effort Game.