Plus/Minus Grading and Motivation:
An Empirical Study of Student Choice and Performance

James E. McClure
Department of Economics
Ball State University
Muncie, IN 47306
jmcclure@bsu.edu

Lee C. Spector
Department of Economics
Ball State University
Muncie, IN 47306
00lcspector@bsu.edu

An Empirical Study of Student Choice and Performance

ABSTRACT

This paper considers whether student motivation might be impacted by the replacement of a straight (A, B, C, D, F) grading system with a plus/minus system (A, A-, B+ . . . D-, F). The data that are examined are from several undergraduate economics classes at a mid-sized Midwestern university in the United States. The data includes student characteristics, student performance, and students' choices of either a plus/minus or a straight grading system. In this admittedly small scale study students, who chose plus/minus grading, were not significantly more motivated than students who opted for straight grading.

Plus/Minus Grading and Motivation: A Study of Student Choice and Performance

I. Introduction

Grades are generally assigned by either a *plus/minus* system (e.g., *A, A-, B+, B, B-, . . .*) or a *straight* system (e.g., *A, B, C, D, F*). A growing number of American universities are switching from straight to plus/minus grading systems.[1] In part, such changeovers have been supported by a first generation of articles and publications that contend that plus/minus systems improve student motivation because: 1) the smaller gaps between grades hold open the possibility of a student being able to improve his/her grade for a longer fraction of the semester; and 2) they appeal to students' sense of fairness.[2] While this first generation of research has proven valuable in placing the motivation aspect of the plus/minus system at the forefront, it doesn't present a rigorous empirical test of the asserted higher motivational effect.

Employing a unique data set, this paper conducts a standard statistical examination of the hypothesis that plus/minus grading gives students superior motivation relative to straight grading. Our main innovation is that the students in our study chose, for themselves, whether to be graded under a plus/minus grading system or under a straight system.

In what follows, we first discuss the following aspects of the literature: a) the non-rigorous arguments that have been directly used to attest to the motivational consequences of plus/minus grading; and b) the types of biases that are inherent in papers that study plus/minus grading. Next, we describe our empirical methodology, model, and results. Finally we summarize the paper's results.

II. Aspects of the Literature

One reason for the popularity of the motivational argument seems to be based upon a series of persuasive hypothetical examples that have arisen out of discussions about alternative grading systems. The following are illustrative:

1.) Average Angie performed reasonably well in her classes. Although she wanted to do better [,] she seemed to get an average C on most exams and assignments. Like most students, Angie is an adept bookkeeper and constantly calculates her class grade given the evaluation criteria in the syllabus. As the semester progresses, Angie realizes that given the assignments yet to be completed, it is mathematically impossible for her to raise her average enough to earn a 'B.' While disappointed, Angie also realizes that it is also nearly impossible for her average to drop low enough to earn a 'D.' Angie's conclusion is that she doesn't need to work any harder since her efforts are unlikely to result in a better grade. More importantly, Angie realizes that she can also slack off without any serious implications to her grade. (Andrew Bessettes, 2002, p. 37)

2.) Zia and Rudy have chosen the same major and are enrolled in that discipline's core sequence of courses. In the first term, Zia tries hard and amasses 445 points that translate to a B grade while Rudy is not confident, noncommittal, and amasses 405

[1] On trends toward plus/minus grading, see Bressette (2002), Quann (1987), and Riley, Checca, Singer, and Worthington (1996).

[2] The fairness argument runs parallel to the motivation argument. When the reward for a 70% course performance is the same as for a performance of 79%, students may think this "unfair" and become discouraged about learning more than 70% of the material if they don't think they can learn more than 79% of the material.

> points that also translate to a B grade. In the second term, Zia is distracted by things extra-curricular and slips 40 points (10% of the possible scale) but still receives a B. Meanwhile Rudy screws up his courage, tries hard, and raises his point total to 445 but still gets a B. In anticipation of a third term, the message received by Zia is likely to be "I can coast since I didn't even try this term and still got a B." At the same time, the message received by Rudy is likely to be "I'm not good enough. I worked a lot harder but it didn't make any difference since I still got a B. I'll never be able to get an A." (Academic Senate for California Community Colleges, 1996, p. 8)

Anecdotal examples, despite their persuasive appeal, do not constitute a test, nor do they establish the motivational advantages of *plus/minus* grading in an empirically rigorous manner.

Another argument about the motivational advantages of plus/minus grading simply asserts them as an extension of a literature that shows that the presence of grades (versus their absence entirely) tends "to support student motivation and success." (Malone, Nelson, and Van Nelson, 2002, p. 11).[3] But given the differences in the orders of magnitude when changing from no grades to grades versus changing from straight grading to plus/minus grading, this seems to be a speculation that begs for rigorous verification. Yet we found no such evidence in the literature.

The paucity of rigorous empirical predictions about the impact of plus/minus grading systems is possibly because such testing is not an easy task.[4] Three biases are present in the existing work in this area. First, difficulties arise because of biases that are inherent in the traditional methods of examining students across classes. If different instructors are involved, there is the possibility that professorial behavior will differ with respect to grading curves and exam corrections when a professor uses a straight grading system as opposed to a plus/minus grading system. For example, one of the implications of Stancato and Eiszler's (1983) work is that professors would be less inclined to give B-'s or C+'s since these grades are the ones that cause students the most psychological problems and would lead to costly grade appeals. To the extent that the degree of risk aversion varies across instructors, this introduces a bias to studies of the course grades resulting from plus/minus systems whenever more than one professor is under study.[5]

Second, the abrupt introduction of a new grading system creates winners and losers within the student body. Students who prefer the old system are likely to behave differently to this change than those who prefer the new one. This suggests that unless these behavioral changes are taken into account, biases may be present in grading system experiments that follow the traditional methodology of comparing a control class (where students continue to be graded by the status quo system) to an experimental class (where all students are graded by the new

---

[3] Some of the examples cited by Malone, Nelson and Van Nelson are Ebel (1974), Eiszler (1983), and Stallings and Leslie (1970).

[4] See Baker and Bates' (1999) literature review for support of this statement.

[5] In the economics education literature, the standard technique for overcoming this type of bias is to gauge achievement by student performance on the nationally standardized TUCE exam (Test of Understanding in College Economics). Although TUCE performance has been widely used as a proxy for student achievement in economics, it is clearly a compromise that is not without some severe shortcomings. On the rationale for using TUCE performance as an acceptable compromise, see Emerson and Taylor (2004). On the biases introduced by using TUCE performance as a measure of achievement, see Becker (1997).

system).[6] On this point, in a study of reactions to the imposition of plus/minus grading, Baker and Bates (1999) found that 59.7% of surveyed students rated its value to them as "negative" (p. 28).

Third, in survey research the question of external validity always arises as to whether respondents would actually choose in a real situation what they say they prefer in a hypothetical situation. For example, while Baker and Bates survey indicates that 59.7% of their students prefer the straight grading system, it does not necessarily mean that if students were actually given the choice of grading systems for themselves, 59.7% would actually choose the straight grading system. This is because there are no sanctions involved with surveys, while, at the same time, survey answers are often given without much thought.[7]

III. Empirical Methods, Model, and Results

Our empirical methods guard against these biases that plague studies of grading. Our key innovation, in this regard, is to empower students with the ability to make the choice for ***themselves*** at the beginning of the semester whether they would, in fact, be graded via either a straight or plus/minus system. In other words, each student was graded under the system he/she preferred. Potential decision biases were overcome by giving students a week to make their choices. This procedure provided students with enough time to carefully consider the matter, and biases introduced by spur of the moment responses were minimized. Additionally, the possibility of any professorial bias was eliminated by having just one professor involved in the teaching of all courses under study. Finally, the professor did not know which grading system each student had chosen until after the final examinations were graded and a scale was formulated. [8]

In order to test for the existence of any motivational benefits to plus/minus grading, we compiled data on student characteristics, achievement, and choices of grading systems (plus/minus or straight). Our strategy is to estimate the impact of student choice of grading system on achievement, after controlling for other student characteristics that are deemed likely to be influential. If the plus/minus system provides greater motivation, we would expect to see higher achievement among students who opt for the plus/minus system (after controlling for other factors that influence achievement).

Our sample included a total of 163 undergraduate students at a Midwestern university from four sections of the Macroeconomics Principles Courses (147 students) and one section of the Intermediate Macroeconomics Course (16 students).[9] Because all of these students had

---

[6] Furthermore, for the same reason, one cannot recalculate, after the fact, what a given class's grade point would have been if it had been graded by a different grading system.

[7] In a survey of a different set of students, we asked our students which grading system they preferred. Two weeks later, they were allowed to actually choose the system under which they would be graded. Twenty-five percent responded differently than on the original survey. While this might be class specific, it does suggest the potential problem of external validity as it relates to students' preferences with respect to grading systems as indicated by survey data.

[8] Additionally, because just one professor was involved, we were not obliged to make the compromise of using TUCE scores to proxy student achievement (see footnote 3 for more on the TUCE examination).

[9] A variety of different aspects of plus/minus grading among graduate students at the same university were studied by Malone, Nelson, and Van Nelson (2002). One important finding of theirs (p. 15) was that plus/minus grading did not lead to grade inflation.

previously taken a Microeconomics Principle Course, their choices (between grading systems) were enlightened by that experience. That is, they had at least a semester's worth of experience regarding the expectations and demands of college professors, in general, and of economics professors, in particular. None of the students in our data set were first semester freshmen and the majority of the students were in their sophomore, junior or senior years. The median level of experience in our sample was that of a sophomore. That is, the students were familiar with how a collegiate economics course stacked up against their capabilities, time constraints, and levels of interest. Additionally, students were informed at the courses' outsets of the percentage cut-offs for the possible course letter grades under either plus/minus or straight grading. Therefore, we consider the choices made by the students in this study to be relatively *informed* choices (as opposed to completely blind) choices.[10] The Intermediate Macroeconomics students were chosen to determine if upper level students reacted differently than beginning students.

All courses under study had the same instructor, thereby eliminating any variation that might be introduced by an instructor's idiosyncrasies. Additionally, this instructor had over thirty years of experience teaching these courses. Throughout these years, the instructor received student evaluations that were consistently excellent and he was given special recognition by his peers on several occasions for excellent teaching. To the extent that accurate assessment of student learning and instructor experience/quality are positively related, this potential bias was reduced in our data. Finally, the instructor gave each class three exams per semester that each featured both multiple choice questions and essay questions (equally weighted).[11] This reduced the well-known potential assessment biases introduced by exclusively relying upon either type of exam question.

In what follows, a variable measuring student achievement will be the dependent variable in an ordinary least squares regression analysis. Of course, there is no perfect way to measure student achievement. Choosing the best proxy possible given our data, we represented student achievement by a vector whose elements are the percentages of total course points earned by the individual students. We denoted this vector as **ACH**.[12] Although it is true these percentages might not perfectly measure how much the students have learned, it is crucial to put this potential problem in the context of the ordinary least squares estimation model that we are using. It is well known that measurement errors in dependent variables only impact regression results, in general, and the reliability of their t-statistics, in particular, when measurement errors are correlated with the independent variables.[13] Because there is no reason to suspect that such correlations are present in our study, imperfections in our measure of student achievement do not

---

[10] That is, the implicit assumption that we are making in this paper that students were capable of making reasonably informed and accurate choices is not a wholly unreasonable one.

[11] There were two exams during the term and a comprehensive final. The two exams were equally weighted. The final exam constituted 50% of the final course grade. Students knew of this weighting procedure from the outset of the course. This again, served to make student choices relatively informed rather than blind choices.

[12] As noted by one anonymous referee, this measure assumes that "a mark of 79% equates to 'learning 79% of the material'." Yet this measure is clearly superior to using a letter grade as a proxy because, as the other referee pointed out, a letter grading measure can magnify the unreliability of an underlying examination. On this score see Please (1971) and Wood (1991).

[13] For a discussion of measurement errors with respect to the dependent variable in a regression, see Greene (2000, pp.375-76) and Pindyck and Rubenfeld (1998; pp. 180-81)

undermine the validity of the results of our statistical analysis.[14]

The novel explanatory variable for achievement in this study is student choice of grading systems. More specifically, we compiled a vector, **CHOICE,** whose elements equal either one for students who chose plus/minus grading or zero for those who chose straight grading. Along with **CHOICE**, traditional explanatory variables of achievement were also taken into account. The following vectors of control characteristics were compiled: 1) **CLS**, whose elements indicate each student's class standing; 2) **GEN**, whose elements designate each student's gender (**GEN** equals one for males and zero for females); 3) **GPA**, whose elements are the cumulative grade point average of each student; and 4) **INTER**, whose elements equal either one for students in Intermediate Macroeconomics and zero for students in Macroeconomic Principles.

Before analyzing the data, students with "A range" GPAs (i.e., above 3.66) were eliminated from the sample. In the plus/minus system under study, a grade of A+ does not exist. The absence of this possibility can only dull achievement incentives at the upper end of the distribution. An "A range" student, who chooses plus/minus grading, will be unrewarded (in terms of her awarded grade) for performing at the A+ level (again, in our school's plus/minus system no A+ grades can be given). Also eliminated from the sample were students for whom no cumulative GPA data existed. This occurred because a student's university record was incomplete.

Table 1 provides an overview of the number of students in various categories and how frequently the plus/minus grading system was chosen within the categories. Comparing the percentages by categories, with the percentage for all students (in the first row), there are only a few striking deviations from the all student percentage of 40.7%. While only 25% of "seniors" and only 30% of Intermediate Macroeconomics students chose plus/minus grading, the sample sizes in these categories are too small to get very excited about these deviations. This said, the lesson of Table 1 is that pair-wise comparisons of the percentages in the various categories with the all-student percentage are suggestive of very little.

Turning from pair-wise comparisons, we proceed to multivariate analysis. More specifically, we examine the proposition that a plus/minus grading system provides superior motivation by estimating an equation of the following form:

$$\mathbf{ACH} = c + b_1\mathbf{CLS} + b_2\mathbf{GEN} + b_3\,\mathbf{GPA} + b_4\mathbf{INTER} + b_5\mathbf{CHOICE} + \mathbf{e},$$

where c is a constant, b_i 's are slope parameters (i=1, . . . 5), and **e** is a random error vector. The results of running this equation for the remaining sample suggest that the choice of plus/minus grading had no significant impact on student achievement. These results are found in Table 2. The key result is that t-statistic for **CHOICE** indicates statistical insignificance. That is, for the students in our sample, plus/minus grading provided no significantly different learning incentives than did the straight grading system. This finding is contrary to the conventional wisdom of many educators who, as detailed in this paper's introduction, hold that plus/minus provides superior motivation over straight grading.

Given the controversial nature of this papers main result, it is important to re-emphasize that our analysis was not without a number of caveats. Recall, for example, that our data

[14] As noted earlier, there are also qualitative reasons to think that assessment reliability was unusually high in our sample owing to: 1) the experience and unusually high quality of the instructor; and 2) the use of both multiple choice and essay questions on exams.

represent only the students at one university enrolled in undergraduate economics courses under the instruction of one professor. In the context of the caveats to our analysis, the strongest conclusion that we would hazard to make is this: The motivational advantages of plus/minus grading over straight grading appear at this point to be rather over-inflated. Furthermore, as economists, we would also offer the following thought. To the extent that plus/minus systems are more costly to administer, further research (in different settings, subject areas, *et cetera)* on the motivational consequences of the plus/minus grading system appears advisable before eliminating less costly straight grading systems.[15]

IV. Summary and Implications

It is generally accepted that motivation is an important key to student success. This is obvious in light of the broad range of issues in education that are influenced by student motivation. For example, studies concerning grade inflation are biased whenever variations in student motivation are embedded in the data, but not controlled for statistically.

In this paper, we explored one aspect of motivation – the impact of the grading system being used. Our goal was to empirically investigate the commonly held notion that plus/minus grading is superior to straight grading in motivating student achievement. This is a timely issue because many universities have recently switched from a straight grading system to a plus/minus grading system, in part, due to the contention that it would have a positive impact on student motivation. This contention is at odds with the main result of this paper: for undergraduates enrolled in a limited number of courses at a mid-sized Midwestern U.S. university, the choice of plus/minus grading had no statistically significant influence upon the percentage of total points earn during a semester. Given the limitations of the data set employed and given that our dependent variable is only a proxy for motivation, one would be premature to take this result as definitive. However, it does argue in favor of the need for further research regarding the motivational consequences of alternative grading systems.[16]

[15] Additionally, as an anonymous referee pointed out, any inferences about achievement and motivation would surely benefit from additional qualitative data on students' attitudes, biases and perspectives on straight and plus/minus grading systems.

[16] This is especially true since our findings seem to contradict the generally held idea that the plus/minus grading system is more motivational than the straight grading system.

References

Academic Senate for California Community Colleges (1996) *Plus and minus grading options: Toward accurate student performance evaluations.*

Baker, H. Eugene III and Homer L. Bates (1999) Student and faculty perceptions of the impact of plus/minus grading: A management department perspective, *Journal of Excellence in College Teaching*, 23-33.

Becker, William (1997) Teaching economics to undergraduates, *Journal of Economic Literature* **35**, 1347-73.

Bressette, Andrew (2002) Arguments for plus/minus grading: A case study, *Educational Research Quarterly* **25** (3), 29-41.

Ebel, R.L. (1974) Shall we get rid of grades? *Measurement in Education* **5**, 1-6.

Eiszler, C.F. (1983) The meaning of college grades in three grading systems, *Educational* Research Quarterly **8** (3), 14-19.

Emerson, Tisha, L.N. and Beck A. Taylor (2004) Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course, *Southern Journal of Economics* **70** (3), 672-693.

Greene, William H. (2000) *Econometric Analysis* (4th edition), Upper Saddle River: Prentice Hall.

Malone, Bobby G., Jacquelyn S. Nelson, and C. Van Nelson (2002) A study of the effect of the implementation of the plus/minus grading system on graduate student grades," *College and University* (Winter), 11-21.

Pindyk, Robert S., and Daniel Rubenfeld (1998*) Econometric Models and Economic Forecasts* (4th edition), Boston: Irwin McGraw-Hill.

Please, N.W. (1971) Estimation of the proportion of examination candidates who are wrongly graded, *British Journal of Mathematical and Statistical Psychology* **24**, 230-38.

Quann, C.J. (1987) Plus-minus grading: A case study and national implications, *American Association of Collegiate Registrars and Admissions Officers*, Washington, D.C., 1-76.

Riley, H.J., R.C. Checca, T.S. Singer, and D.E. Worthington (1996) Current trends in grades and grading practices in undergraduate higher education: The results of the 1992

AACRAO survey, Washington, D.C.: American Association of Collegiate Registrars and Admissions Officers.

Stallings, W.M., and E.K. Leslie (1970) Student attitudes toward grades and grading, *Improving College and University Teaching* **18**, 66-68.

Stancato, Frank A., and Charles F. Eiszler (1983) When a 'C' is not a 'C': The psychological meaning of grades in educational psychology, *Journal of Instructional Psychology* **10** (3), 158-162.

Wood, R. (1991) *Assessment and Testing: A Survey of Research*. Cambridge: Cambridge University Press.

Table 1 – Summary Statistics

	Number of Students	Percentage Choosing Plus/Minus
Entire Sample of Students	135	40.7%
Students with GPA > 2.66	59	42.4%
Students with GPA < 2.66	76	39.4%
Male Students	87	39.0%
Female Students	48	44.0%
Intermediate Macro Students	10	30.0%
Freshmen	51	45.0%
Sophomores	50	38.0%
Juniors	20	45.0%
Seniors	12	25.0%

TABLE 2 – OLS Regression

Dependent Variable: ACH
Method: Least Squares
Sample: 1 163: IF GPA < 3.66
Included observations: 137
Excluded observations: 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.20	4.02	8.99	0.000
CLS	- 0.38	0.89	-0.43	0.670
GEN	1.36	1.85	0.73	0.465
GPA	13.03	1.25	10.35	0.000
INTER	- 7.06	3.89	-1.81	0.072
CHOICE	- 1.61	1.76	-0.91	0.364

R-squared	0.456	Mean dependent var	68.80
Adjusted R-squared	0.435	S.D. dependent var	13.407
S.E. of regression	10.076	Akaike info criterion	7.501
Sum squared resid	13301.84	Schwarz criterion	7.629
Log likelihood	-507.82	F-statistic	21.950
Durbin-Watson stat	1.695	Prob(F-statistic)	0.0000