

Is Finer Better?

A Master's Investigation into Coarse Grading

Andrew K. Rose*

March 9, 2022

Abstract

I test the hypothesis that a typical grading scheme with ten passing grades (A+, A, A-, B+, ...) contains no more information than a coarser scheme with only three passing grades (distinction, merit, and pass). My empirics exploit cross-sectional data from the individual grades of over 1200 master's students who took more than 200 courses at the National University of Singapore's Business School during the 2019-20 academic year. I find little evidence that a fine grading scheme is more informative than a coarse scheme. Survey evidence, from students and staff, collected before and after a 2021 switch from fine to coarse grading at NUS-Business show little evidence that the change in grading policy affected student academic conduct. If coarse grading schemes have few adverse consequences for motivation or information, but encourage teamwork, experiential learning, and/or extra-curricular activities, they may improve the welfare of students, faculty, and grade-users like employers.

Keywords: empirical; data; policy; business; school; survey; scheme; course; module; assessment.

Contact: bizaro@nus.edu.sg; +65 6516-3075; NUS Business School, 15 Kent Ridge Dr., Singapore 119245.

* Distinguished professor and dean, NUS Business School. I thank: Helen Chai and Ivan Png for comments; Huiyu Yeo and participants for providing survey data; task force and AMM participants at NUS-Business for feedback. All errors are mine alone. Key output and a copy of this working paper are freely available from <https://bschool.nus.edu.sg/dean/research/>; the (confidential) grading data set cannot be posted, but if you send me an appropriate program (check the posted output!), I'll run it for you and send you the output, at least for a reasonable period of time after release, so long as it doesn't reveal confidential information (or waste too much of my time).

Introduction and Motivation

The key purpose of student assessment is to identify what students have learned, to a) determine if they have mastered the material in an absolute sense, and b) evaluate their understanding relative to their current and past peers. In doing so, the assessment provides incentives for students to learn the material. Unfortunately, student assessment can also, inadvertently, create distortions. For instance, some students can become fixated on activities directed towards grades to the detriment of learning both inside and outside the classroom.

In this paper, I ask: “Is a granular grading system more informative than one that is coarser?” I compare a typical grading system with a large number of passing grades (A+, A, A-, B+, ...) to one with a small number. Using recent data from over twelve-hundred master’s level students at the National University of Singapore (NUS) Business School, I present evidence that little information is lost when instructors use three coarse rather than ten fine passing grades. NUS-Business switched from fine to coarse grading for these students in 2021, and survey evidence from before and after the change confirms that student academic conduct changed little, as perceived by both students and staff.

The Larger Context

Most of this article is devoted to a narrow empirical question, namely “How informative is a fine grading policy compared to a coarse scheme?” Still, the setting for this investigation is also relevant since grading policy has several effects.

A good grading system appropriately summarizes students’ absolute and relative understanding of the material, while minimizing other consequences. At the master’s level in a business school (the particular focus of this article), such a system should ideally:

1. Encourage appropriate behavior; part of the motivation for student academic engagement is the incentive provided by grading, but grading can also lead to excess focus on grades,
2. Encourage students to be adventurous, particularly at the master’s level where students are interested in considering a new or unexplored field, industry, or occupation,
3. Encourage teamwork, and facilitate experiential learning,

4. Encourage suitable extra-curricular engagement and leadership, since club/organization activities and networking are important parts of the masters-level business experience and are often highly valued by recruiters as well as students,
5. Be transparent, equitable and just, so that students can reasonably expect faculty to adhere to the policy, and
6. Contain stress to reasonable levels in both faculty and students, by encouraging a focus on learning rather than assessment.

There are not the only desirable attributes of a grading system. Green and Emerson (2007) also include: a) transparency; b) ease of implementation; c) ease of aggregation across courses; d) reliability across graders; e) stability over time; f) horizontal consistency across students, so that similar achievement receives a similar graded; and g) the ability to differentiate vertically among different levels of achievement so that different achievement receives a different grade.

Changing grading policy is difficult. Not only do policy changes have to be well motivated and explained, but ideally, they should also:

1. Be easy to explain to and popular with (most) potential students and employers,
2. Be plausible and defensible (a good pedigree helps),
3. Not reduce academic effort substantively,
4. Not adversely affect (m)any students whose tuition is paid by their employers, and
5. Not adversely the employment prospects of (m)any students.

The motivation for this analysis is not strictly relevant to understand what follows, but it may be helpful to understand the institutional setting. Grades are an obsession for (some) master's students, at least at NUS-Business. Since some students are fixated on grades, faculty must spend considerable time on fine gradations. All this may be a distraction from the larger experience for students, and wasteful for our faculty members. The system may also inadvertently discourage teamwork, experiential, and extra-curricular activities, and add stress. More studious students often have weaker life-, soft- and people-skills. Further, university graduates who have spent disproportionate time on bookish activities may be less creative, even if they perform better at well-defined tasks. Thus, there is, at least in principle, the potential for a different grading policy to help, at least a little, with some of these problems.¹ A

key concern though is whether a fine grading scheme contains significantly more information about student achievement than a coarse scheme, and thus might provide less motivation for academic effort. The potential loss of such information would constitute a serious argument against any policy switch; this concern provides the motivation for the analysis that follows.

Objective, and Preliminaries

The goal of this paper is to argue, based on data, that a simpler grading policy would have negligible effect in practice on the informativeness of the grading system. I show that there is little evidence that abandoning a system with ten passing grades (A+, A, A-, B+, ...) for one with only three (Distinction, Merit, Pass) would distort the information value of the grading system. Such a system might also have other benefits – for these, I necessarily have no data yet – so a simpler system might be a net win.

My argument rests primarily on analysis of data from the population of master's level students at NUS Business School from the 2019-20 academic year. I have grades for each of the >1200 students for each of their >200 modules, all operated under a fine grading system; I compare these to a set of hypothetical coarse grades. Since NUS-Business changed its masters-level grading policy from fine to coarse in 2021, I also use survey data, collected from both students and staff in the semesters immediately before and after the change in policy. Neither data set indicate any substantive loss from a switch to coarse grading.²

A small academic literature on this issue examines both the theory and empirics associated with the granularity of the grading system; it is covered briefly in an appendix. I summarize the extant theoretical literature as indicating that a coarse system can elicit either more or less student effort than a more granular one. The literature provides a little weak empirical evidence that a coarser system has effects in practice and considerable evidence that it does not. Given the ambiguous theoretical literature and the paucity of relevant evidence, the objective of this research is wholly empirical in nature. The goal of this paper is to examine and compare the amount of information conveyed by fine and coarse grading schemes.

The Empirical Setting

Grading Policy for NUS-Business

NUS-Business, as a typical unit within the National University of Singapore, has historically used a conventional grading system for its master’s students, presented in Table 1. There are ten passing grades that vary from A+, A, and A-, ..., through D as well as F (failure).³ For convenience below, I refer to this as a “ten-cell” system. Almost every course – at NUS, referred to as a “module” – taken by a student is assigned a letter grade, derived from underlying numerical “marks”.⁴ Grades are assigned numerical grade points, which are also presented at the top of Table 1. At NUS, these grade points vary from 1 through 5.⁵ The average of grade points, weighted by the number of units (“module credits” at NUS) constitutes the “Cumulative Average Point” or “CAP”.⁶ Also included in Table 1 is the percentage of grades assigned for modules taken by NUS-Business master’s students in 2019-20, using the ≈7000 module-student grades.

Table 1: NUS-Business Fine Grading Scheme Mapped to a Hypothetical Three-cell Scheme

Status Quo	Old (Fine) Grades	A+	A	A-	B+	B	B-	C+	C	D+	D	F
	Grade Points	5	5	4.5	4	3.5	3	2.5	2	1.5	1	0
	Percentage	6.9	18.4	22.8	23.7	17.0	7.5	3.0	0.6	0.0	0.2	0
Rough Hypothetical	New (Coarse) Grades	Distinction			Merit			Pass				F
	Grade Points	5			4			3				0
	Percentage	25.3			63.6			11.3				0

Data from all modules taken 2019-20 by master’s students at NUS-Business.

An Alternative Three-Cell System

Harvard Business School uses the following grading policy:

- Category I “Given to the top 15%-20% in [required] or the top 15%-25% in [elective] courses.”

- Category II “Given to the next 70%-75% in [required] or the next 65%-75% in [elective] courses ... the actual number is subject to the number of Category I grades assigned.”
- Category III “Given to the lowest-performing 10% of students.”⁷

A three-cell system like this allows the instructor to: a) encourage and reward excellence; b) deter and discourage poor performance, and c) encourage most other students to fixate on learning and the student experience, not their grades. Beyond the three passing cells, all grading policies must allow students to fail, so another cell exists for failure (rarely used at the master’s level).⁸

While HBS’s is not the only coarse grading system in use, I use it to compare with old NUS-Business master’s practice. It is a well-defined alternative from a reputable institution. Further, in January 2021, NUS-Business school voted to move away from its ten-cell system to Harvard’s three-cell system at the master’s level, albeit with different labels for the cells (Distinction, Merit, and Pass in place of Categories I, II and III respectively). Now, NUS-Business instructors continue to collect numerical marks in precisely the same way as they did when there were ten passing grades; the marks are simply mapped onto a smaller number of coarse grades.

A rough mapping from the (old) NUS-Business grading system to a hypothetical three-cell scheme is included in the bottom half of Table 1. This is done by simply substituting: Distinction for A+ and A grades; Merit for A-, B+ and B grades; and Pass for all passing grades of B- and below.^{9,10}

Data from NUS-Business

I now examine historical master’s grading patterns from NUS-Business, using data from its previous ten-cell policy. My goal is to compare these actual grades with what might have happened if the data had been collected in the same way but replaced with the (new) three-cell set of coarse grades.

An Illustrative Example

The essence of a coarse grading system is best conveyed with an illustrative example. Figure 1 contains histograms from an actual and typical master's module taught a couple years ago at NUS-Business. The final marks received by each of the 74 students at the semester's end are displayed in a histogram; these are numerical and range from 50.3 through 83. The instructor grouped these into six grades, ranging from "C" through "A"; these histograms are clearly marked with dashed horizontal lines.

Figure 1 implicitly shows the dividing lines between fine official grades (A, A-, ...), as well as the corresponding cutoffs for synthetic coarse grades (Distinction, Merit, ...), also shown with thick solid horizontal histogram lines. These dividing lines cause much angst in both students and faculty; the former try to ensure that they are on the right side of the border, while the latter must create some sort of divide, no matter how tenuous. For instance, B+ was awarded to 21 students in this module; their marks ranged from 70.2 to 74.4. But the two students with marks of 74.4 could reasonably argue that each was essentially observationally equivalent to the A- student with 75 marks. Such narrow and somewhat arbitrary dividing lines are inevitable and unavoidable; they could become more important with coarse grades. But students and faculty may spend less time obsessing about them if there are: a) fewer dividing lines, that are b) quantitative, and c) transparent (the last important for handling students close to a margin).

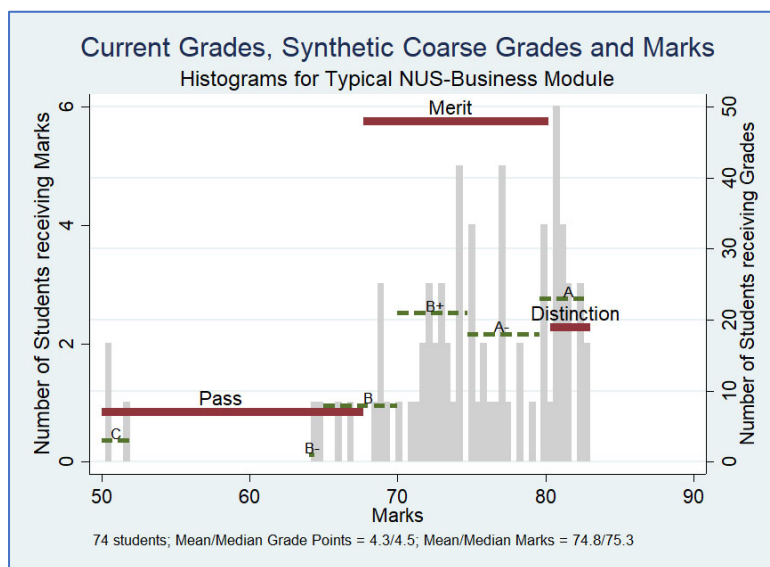


Figure 1

Data for the same module are portrayed in Figure 2 but via the cumulative distribution function instead of histograms. Above the CDF are triangles corresponding to fine letter grades; the height indicates the proportion of students receiving the grade, while the breadth indicates the range of marks. Below the CDF are analogous triangles for the synthetic coarse grades.

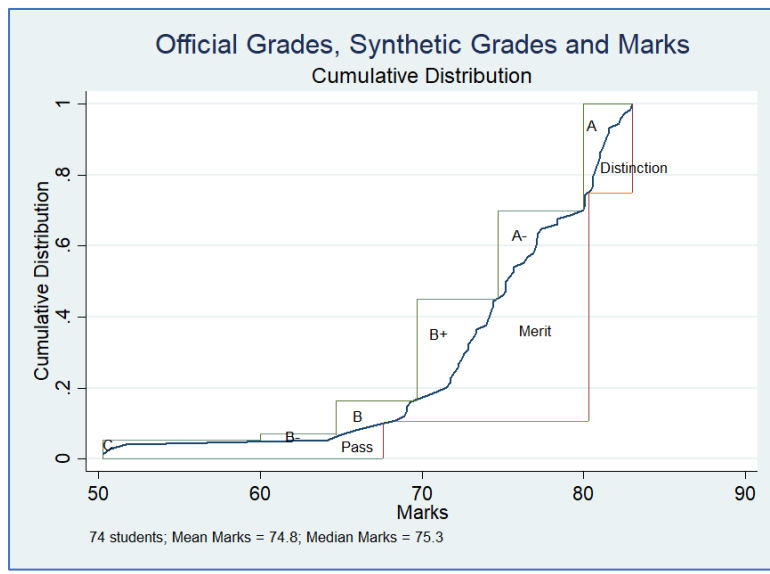


Figure 2

I now begin to analyze the data set at a more aggregate level.

What do the Grade Distributions Look Like?

For NUS-Business as a whole, only seven of the ten passing grades were really used in practice, C+ through A+; less than 1% of grades are below. Histograms of all grades earned by students in all NUS-Business masters-level modules in AY 2019-20 are displayed in the top-left panel of Figure 3. Analogues for the three distinct types of master's programs are similar and displayed in other panels.¹¹

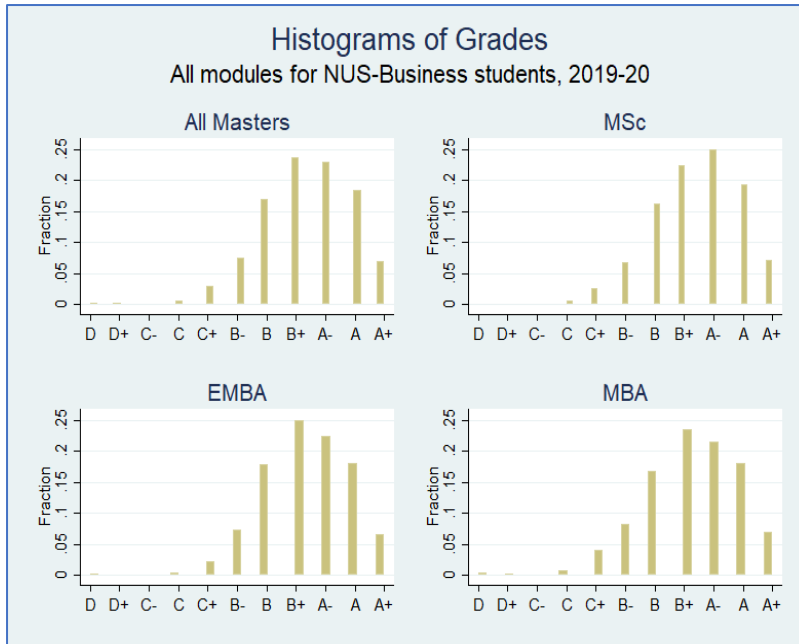


Figure 3

It's also revealing to present the aggregate data with cumulative distribution functions; I do so in Table 2. I have added extra thick lines at approximately the points where a three-cell scheme would place them.¹²

Table 2: NUS-Business Grade Cumulative Distributions

All Masters Courses			Cumulative			
Grade	Frequency	Percentage	All	MSc	EMBA	MBA
D	12	0.17	0.17	0.00	0.17	0.32
D+	1	0.01	0.19	0.00	0.17	0.36
C-	0	0.00	0.19	0.00	0.17	0.36
C	43	0.62	0.81	0.62	0.60	1.16
C+	206	2.96	3.77	3.24	2.76	5.17
B-	519	7.47	11.24	9.95	10.03	13.46
B	1,180	16.98	28.22	26.14	27.89	30.28
B+	1,645	23.67	51.88	48.57	52.85	53.74
A-	1,589	22.86	74.75	73.67	75.30	75.13
A	1,278	18.39	93.14	92.95	93.37	93.07
A+	477	6.86	100.00	100.00	100.00	100.00

Data from all modules taken 2019-20 by master's students at NUS-Business.

With a feel for how actual master's grades looked at NUS-Business for 2019-20, it is time to create a hypothetical analogue to these grades for a three-cell system. This enables one to compare actual (fine) with synthetically created (coarse) grades. And with these synthetic grades, it is possible to examine the implications of the grading policy on hypothetical CAPs.

What Would a Three-Cell Scheme Have looked Like Using Historic Data?

I analyze this at three levels of granularity: by module grades, by student, and by module.

Actual and Three-Cell Grades at the Student-Module Level

First, consider the grades received by individual students, one for each module taken. There are approximately 7,000 grades earned by students for modules taken in 2019-20. From this data set, I create hypothetical grade points students would have received under a three-cell system. I do this by using the mapping of Table 1; substituting grade points=3 (Pass) for grades of B- or under, grade points=5 (Distinction) for A and A+, and grade points=4 (Merit) for the others in between.

Actual and hypothetical grade points have similar distributions as shown in Table 3. The distribution under the hypothetical three-cell system is necessarily tighter; since there cannot be any synthetic grade points under 3 by construction, the spread (and the standard deviation) of the data is necessarily lower. But the means and medians line up closely, as do both higher and lower percentiles.¹³

Table 3: Individual Grade Points under Actual (Ten-cell) and Hypothetical Three-cell Grading

Percentile	Actual				If grades assigned under three-cell			
	All	EMBA	MBA	MSc	All	EMBA	MBA	MSc
5%	3	3	2.5	3	3	3	3	3
10%	3	3	3	3.5	3	3	3	4
25%	3.5	3.5	3.5	3.5	4	4	4	4
50% (median)	4	4	4	4.5	4	4	4	4
75%	5	4.5	4.5	5	5	4	4	5
90%	5	5	5	5	5	5	5	5
95%	5	5	5	5	5	5	5	5
Mean	4.14	4.15	4.10	4.19	4.14	4.15	4.11	4.16
Std. Dev.	.72	.69	.76	.69	.59	.57	.61	.58
Observations	6,950	2,352	2,498	2,100	6,950	2,352	2,498	2,100

All NUS-Business master's student-module grades, 2019-20.

Actual and Three-Cell Grades at the Student Level

The next level of granularity is that of the individual student; over 1200 NUS-Business masters-level students took modules in 2019-20. The student-specific CAP is the (weighted) mean of the actual grade points those students received for modules taken in 2019-20. The student's hypothetical CAP is computed analogously, substituting the grade points that they would have received from a hypothetical (coarse) three-cell system in place of the actual (fine) grade points.

Actual and hypothetical CAPs are highly correlated, as shown in Figure 4. The actual (on the x-axis) and synthetic (on the y-axis) student averages are tightly clustered around the (white) 45° line, indicating that there is little loss of information associated with using the three-cell coarse grading scheme instead of the ten-cell system. Peering closely, one can see that the only non-trivial effect is that a few marginal students in the EMBA and MBA programs would have CAPs of 3 or more under the new system, when the fine system gives them 2s.

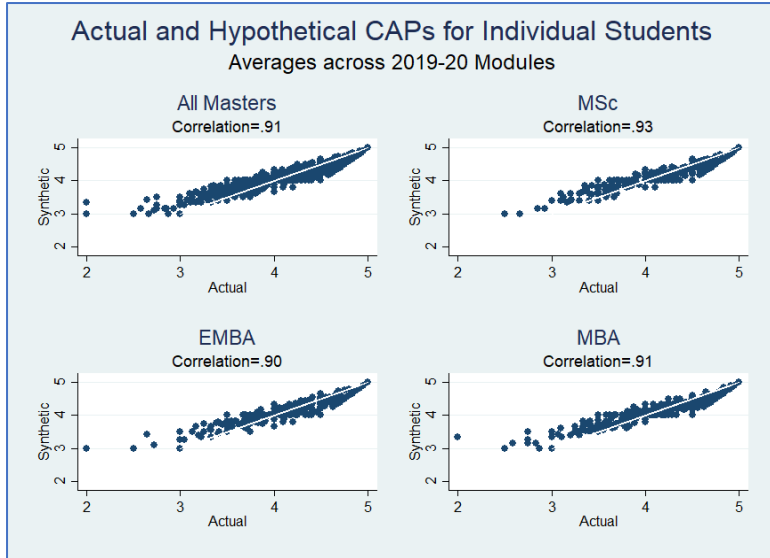


Figure 4

Since many constituencies take CAPs seriously, it is worthwhile to examine the effects of coarse grading on CAPs in a little more detail. Figure 5 displays a histogram of the deviations of a student's official CAP from the CAP that they would have received from the coarse three-cell grading system of Table 1. The effect of using coarse rather than fine grading on student CAPS is small. The deviations are tightly centered around zero; a few negative outliers noted above.¹⁴

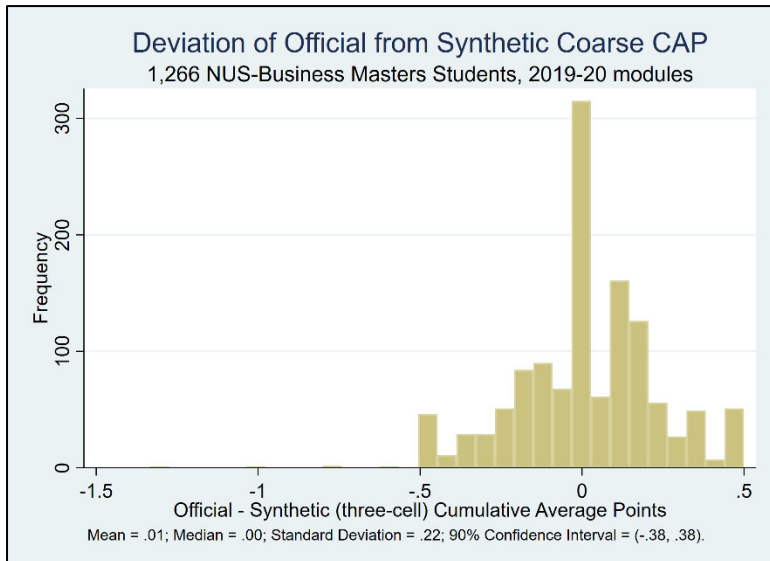


Figure 5

Actual and Three-Cell Grades at the Module Level

The final level of granularity is that of the module; NUS-Business students took over 200 masters-level modules in 2019-20. I compute averages across students, resulting in a single average for each module offered. The module-specific average grade-point is the cross-student mean of the actual grades received; the module's hypothetical analogue is the same thing, substituting the grade points that would have been given from a hypothetical three-cell system for the actual ones received.

Figure 6 shows that actual and hypothetical average grade-points are essentially as highly correlated when one looks across modules as when one looks across students (as in Figure 4). Yet again, the observations are tightly clustered around the 45° line; there are few substantive outliers.

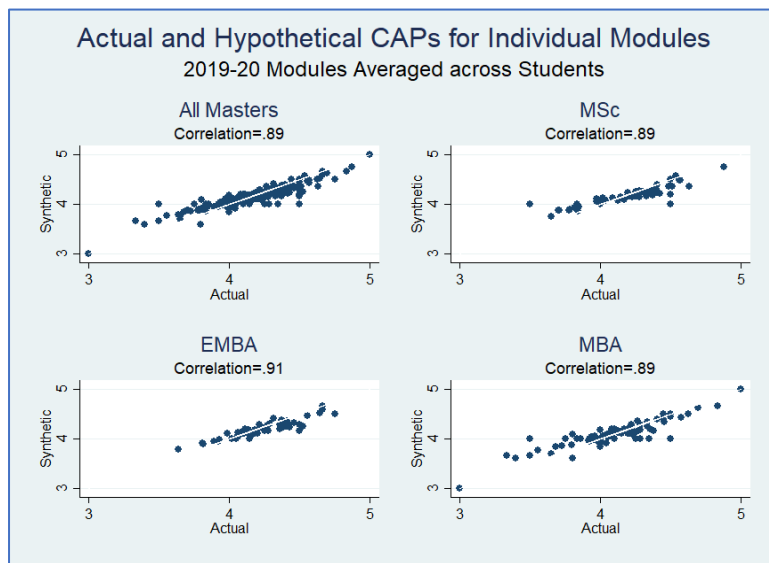


Figure 6

For those who find tables easier to read than graphs, Table 4 provides aspects of the distributions for actual and hypothetical CAPs, averaged (separately) across both students and

modules. As expected, the distributions are somewhat tighter for the synthetic CAPs. But means and medians are closely aligned, as are high and low percentiles.

Table 4: Cumulative Average Grade Points under Actual and Hypothetical Three-Cell Grading Policy

Percentile	(1130) Student Averages		(215) Module Averages	
	Actual	Hypothetical	Actual	Hypothetical
5%	3.3	3.5	3.5	3.8
10%	3.5	3.7	3.8	3.9
25%	3.8	4	4	4
50%	4.2	4.1	4.2	4.1
75%	4.5	4.4	4.4	4.3
90%	4.8	4.7	4.5	4.4
95%	4.9	4.8	4.6	4.5
Mean	4.15	4.13	4.15	4.14
Std. Dev.	.51	.39	.32	.25
Std. Dev.	.51	.39	.32	.25

NUS-Business, 2019-20.

I conclude from all this that the grades of a coarse three-cell grading scheme convey essentially the same information as those from a fine ten-cell grading policy.

Survey Results from the Change in Grading Policy

NUS-Business decided to switch from fine to coarse grading in 2021, partially because of the information provided above. To help gauge the effects of this policy change, all masters-level students were invited to answer a brief survey immediately after the second semester of the 2020-21 academic year, when the policy switch from fine to coarse grading had been passed but not yet implemented. Over 400 students responded.¹⁵ Besides two identifier characteristics (nationality and program), the students were asked for their reactions to three substantive questions:

1. "Grades are the key motivator for students to engage in academic work"
2. "I focus on my grades rather than on learning and the wider university experience"
3. "I spend time disputing grades/marks (e.g., arguing for A- rather than B+)"

Answers were collected using a five-point Likert scale, so a student could strongly disagree, disagree, remain neutral, agree, or strongly agree with each of the three statements.

A comparable number of students answered the same quick survey immediately after the first semester of the 2021-22 academic year, when the policy switch had just been implemented. After grading policy changes, the respondents were also asked to respond to:

4. I am less motivated to engage in academic work by the new (coarse) grading system than I was previously.

Student responses are presented as histograms in Figure 7; the gray bars indicate responses *before* the policy switch, while the outlined hollow bars indicate the responses *after* the policy switch. Analogues for staff members (both academic instructors and program staff) are presented in Figure 8.

What do the survey responses indicate? The primary impression one has from the figures is that the distribution of responses before the policy switch is similar to that afterwards. For students, there was a slight shift in responses to "I focus on grades, not learning ..." This took the form of fewer students disagreeing after the change, and more strongly disagreeing. After the change, more students were neutral on "I dispute grades/marks" and fewer strongly disagreed. But little else changed. Both students and staff continue to agree that grades are a key academic motivator. Most students do not think that they dispute grades/marks much, while staff think that students do; this divergence was unaffected by the policy change. Most students state that they are focused on learning and the wider experience, while most staff think they focus on grades; this too remained unchanged. Overall, it seems like the move to coarse grading changed little. This finding is most dramatically shown by responses to Q4; only a quarter of students stated that they were less motivated to academic work by coarse grading, while only a tenth of staff consider students to be less motivated by coarse grading. A strong majority of both staff and students thought that coarse grading did not reduce student motivation, or had no opinion on the matter. This is consistent with the hypothesis that fine and coarse grading schemes convey similar information, so the choice has little effect on academic conduct.

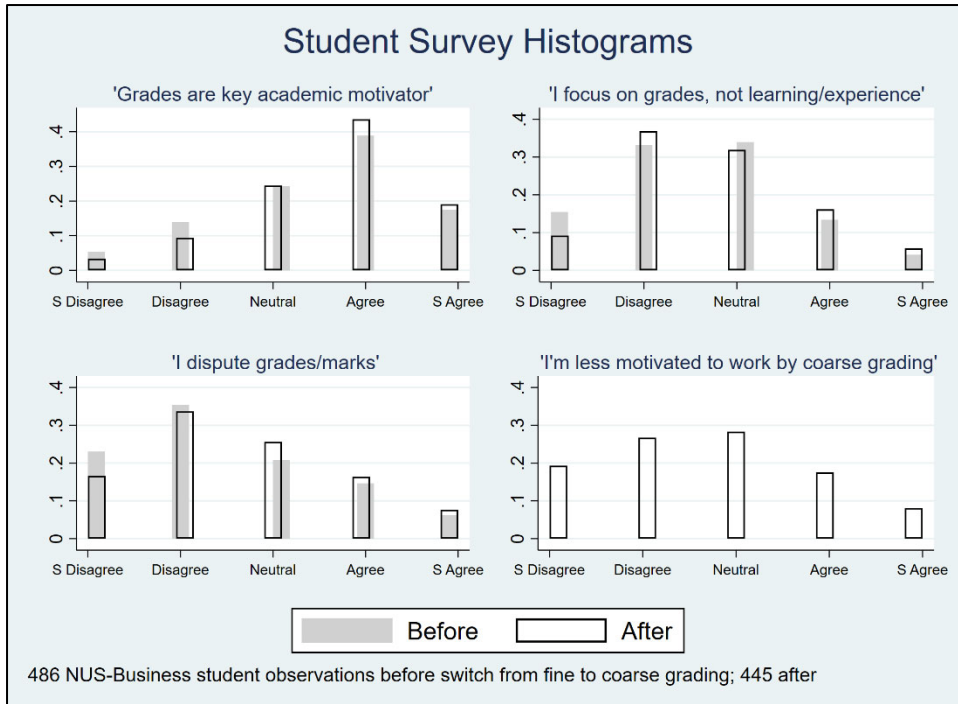


Figure 7

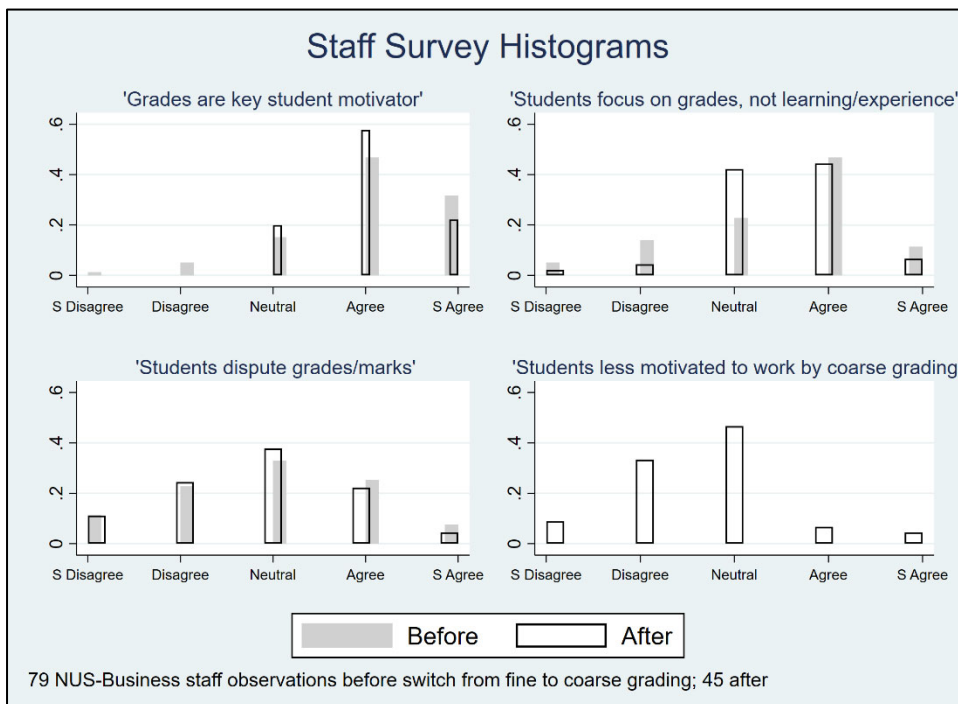


Figure 8

Table 5 provides more rigorous statistical analogues to the visual impression given by the histograms of Figures 7 and 8; it provides chi-square tests of equality of the distribution of responses to each of the questions *before* and *after* the policy switch. The results indicate differences in survey responses to the questions that are minor, insignificant at conventional significance levels. The same is true when one examines subsets of students and staff.

Table 5: Survey Data Tests of Hypothesis of No Effect of Switch from Fine to Coarse Grading

		Obs.	Q1	Q2	Q3
Students	All	931	.10	.03	.07
	MSc	628	.91	.59	.69
	Singaporeans	332	.52	.32	.51
Staff	All	124	.31	.12	.94
	Academic	96	.46	.54	.74

Statistics tabulated are p-values for null hypothesis of no effect of switch from fine to coarse grading, from Chi-Square (4) tests. Questions answered by respondents on five-point Likert scale:

- Q1: "Grades are the key motivator for students to engage in academic work"
- Q2: "I focus on my grades rather than on learning and the wider university experience" (students); "Students spend more time focusing on grades than on learning and the wider university experience" (staff)
- Q3: "I spend time disputing grades/marks (e.g., arguing for A- rather than B+)" (students); "I spend time handling disputed grades/marks with students (e.g., those arguing for A- rather than B+)" (staff)

I conclude that the switch from fine to coarse grading has, at least thus far, had little effect on academic motivation, consistent with only small effects on the informativeness of the grading system.

Conclusion

In 2019-20, a certain master's student at NUS-Business completed seven modules, receiving the following grades: two B (grade points=3.5), two B+ (grade points=4.0), two A- (grade points=4.5), and one A+ (grade points=5.0), thus earning a CAP of 4.14. A three-cell policy would plausibly have led to this student receiving six "merit" (grade points=4.0) and one "distinction" (grade points=5.0), with an average CAP of 4.14. This similarity might seem contrived, but it is not, as shown by the evidence above. Replacing the ten passing grades

previously used by NUS-Business school at the master's level with just three coarse grades results in similar grade averages for both students and modules. That is, little information is lost when a fine grading scheme is replaced with a coarse one.

Any change in grading policy will change the behavior of both students and faculty; indeed, that is the point of a policy change. But, subject to this caveat of endogeneity, there is little statistical evidence that much information should have or has been lost when NUS-Business moved from a ten-cell grading scheme to a three-cell scheme for its masters' programs. A coarse grading scheme plausibly reduces stress for both students and faculty, and encourages faculty to stick to stated policy. It might be expected to encourage marginally more adventurousness and slightly reduce grade-obsession on the part of students. The latter is desirable if it encourages more teamwork, experiential learning, engagement with and leadership in extra-curricular and club activities, and reduced stress. NUS-Business would like to encourage its students to acquire soft-, life- and speaking-skills, as well as academic prowess. A coarser system could reduce effort by students, but it might not, particularly if the stigma associated with a "Pass" is high; the existing academic literature raises few concerns about adverse effects on motivation. The fact that some of the best business schools in the world use a coarse system indicate that the costs can't be that large and are probably negative on net; otherwise, Harvard and the like wouldn't stick with it.

For all these reasons, in early 2021, NUS-Business school decided to switch from a fine to a coarse grading policy. This decision was also driven in part by the finding, presented in this paper, that a coarse grading scheme with just three passing grades has essentially the same information content on a student- or module-basis as that of a fine grading scheme. The survey evidence collected from students and staff members in the semesters immediately before and after the policy change indicate that the move to coarse grading has had few consequences on the academic conduct of students. In years to come, NUS-Business looks forward to investigating data to investigate how this regime change affected other outcomes, such as academic adventurousness, teamwork, experiential learning, and extra-curricular engagement.

References

- Boleslavsky, Raphael and Christopher Cotton (2015) "Grading Standards and Education Quality" *American Economic Journal: Microeconomics* 7(2), 248-279.
- Dubey, Pradeep and John Geanakoplos (2010) "Grading exams: 100, 99, 98, ... or A, B, C?" *Games and Economic Behavior* 69(1), 72-94.
- Grant, Darren and William B. Green (2012) "Grades as Incentives" *Empirical Economics* 44, 1563-1592.
- Green, Kris H. and Allen Emerson (2007) "A New Framework for Grading" *Assessment & Evaluation in Higher Education* (32-4), 495-511.
- Harbaugh, Rick and Eric Rasmusen (2018) "Coarse Grades: Informing the Public by Withholding Information" *American Economic Journal: Microeconomics* 10(1), 210-35.
- Main, Joyce and Ben Ost (2014) "The Impact of Letter Grades on Student Effort, Course Selection, and Major Choice: A Regression-Discontinuity Analysis" *Research in Economic Education* 45(1), 1-10.
- Oettinger, Gerald S. (2002) "The Effect of Nonlinear Incentives on Performance" Evidence from 'Econ 101' " *The Review of Economics and Statistics* 84(3), 509-517.
- Sadler, D. Royce (2005) "Interpretations of Criteria-Based Assessment and Grading in Higher Education" *Assessment & Evaluation in Higher Education* 30(2), 175-194.
- Schinske, Jeffrey, and Kimberly Tanner (2014) "Teaching More by Grading Less (or Differently)" *CBE Life Sciences Education* 13(2), 159-166.
- Starch, D. (1913) "Reliability and Distribution of Grades" *Science* 38, 630-636.
- Wagner, Lilo (2015) "Pass/Fail, A-F, or 0-100? Optimal Grading of Eager Students" DIW Discussion Paper 1445.

Appendix 1: Brief Review of the Academic Literature on Grade Granularity

A Little History

The A-F system of grades has only been around since the late nineteenth century; Schinske and Tanner (2014) provide a brief history. It was first instituted at Harvard and was explicitly intended to diminish motivation, as pointed out by Grant and Green (2012), who quote the 1885 *Annual Report of the President of Harvard* (emphasis added): “The Faculty last year did away with the minute percentage system of marking, and substituted a classification of the students in each course of study in five groups ... It is hoped that this grouping system will afford sufficient criteria for the judicious award of scholarships, honorable mention, and the grade of the bachelor’s degree, *while it diminishes the competition for marks and the importance attached by students to College rank* in comparison with the remoter objects of faithful work ...” The same thing took place in the primary/secondary school system at the same time, in part because precise percentages were viewed as unreliable indicators of achievement. This imprecision arose after evidence was provided a century ago that different teachers were unable to assign grades consistently (Starch, 1913). Grant and Green emphasize that the grading system, particularly the introduction of thresholds, was *not* designed to motivate students (if anything, the opposite).¹⁶

Consistent with this, in 1967 Yale College voted to abolish its numerical grading system and establish a three-cell system (pass, high pass, and honors) precisely to reduce grade obsession. They did so since “it is believed by those members of the faculty who voted in favor of the plan that it will offer some relief from what one educator described as ‘the excessive preoccupation with number of letter grades’ ... ” (quoted in Wagner, 2015).

Theory

In principle, student assessment can be done on a continuous mathematical scale. But, to the best of my knowledge, no student ever receives marks or grades with even three decimal places; instead, for each module they tend to receive one of a small number of integer or letter grades. That is, every institution has implicitly chosen to use a coarse grading system. The question naturally arises “What is the optimal degree of granularity/coarseness?” A small theoretical literature addresses the question.

Dubey and Geanakoplos (2010) consider the matter and conclude that “if students care primarily about their status (relative rank) in class, they are often best motivated to work ... by clumping them into coarse categories. When student abilities are *disparate*, the optimal absolute grading scheme is always coarse ...” Coarse grading often motivates students to work harder because status provides motivation, and coarse hierarchies can (paradoxically) create more competition for status and better incentives to work.¹⁷

Wagner (2015) analyzes the case where students overweight grades, and their behavior thus leads to excessive effort; coarse grading can be used to counterbalance incentives. This is consistent with the motivation of the Harvard/Yale faculty who introduced coarse grades for precisely this reason; for Wagner, “grade obsession induces students to exert inefficiently high effort ... coarse scales can be used to counterbalance incentives ...”

So, coarse grading can, in principle, induce students either to exert higher effort (Dubey and Geanakoplos) or lower effort (Wagner).

There are other reasons why grades are not granular. One associated with information theory is that information may be maximized by a signal with discrete steps, in the presence of measurement noise; coarseness reduces the chance of misclassification. Coarse grades may also be easier to understand to a casual observer (think of hurricane categories). More analysis and references are provided by Boleslavsky and Cotton (2015) and Harbaugh and Rasmussen (2018).

Empirics

Grant and Green (2012) are concerned with measuring the effects of thresholds on student effort, empirically. They use a data set of over 3,000 observations of courses taken by American college students from five instructors at two universities over a decade. Their explicit interest concerns the discrete nature of the A/B/C/D/F grading system, particularly the fact that, in principle, the expected marginal benefit of increased study effort is highly nonmonotonic; the benefit rises then falls in the neighborhood of a grade threshold. However, they find no evidence that students exert more study effort when their pre-exam score is just below the threshold needed to achieve a higher grade, nor is there a bunching of students just above grade thresholds. That is, they find that students who are just below the threshold of earning (say) an A in a module because of a good final exam performance seem to work the same as students who are just above the threshold. This stands in contrast to the results of Oettinger (2002) who used a narrower data set for a single course and found some evidence that student grades clustered slightly above the boundaries separating grades, as predicted by a model with an absolute (not relative) grading standard. Grant and Green (2012) conclude that “stronger incentives are not associated with improved performance...grades are not effective motivators. While this result might be specific to the context, there is no Asian or Singaporean evidence to the contrary, particularly at the masters’ level.

The negative results of Grant and Green (2012) are bolstered by those of Main and Ost (2013) who searched for but found no evidence that numerical scores were clustered around letter grade cutoff thresholds. Where Oettinger (2002) used evidence from a single course, Main and Ost use over 20,000 numerical scores and letter grades from over 9,000 students taking 65 courses in six fields over eleven years.

Appendix 2: Variation across Programs and Course-Type

Variation Across Programs

One could, in principle, have different grading policies for different masters-level programs; the MBA, MSc, and EMBA programs might all have different grading policies at NUS-Business. This would come at the cost of considerable complication, so it is worthwhile to consider whether it would be substantive in effect. Reassuringly, there doesn't seem to be much evidence that it would make a difference, at least based on historical data.

Consider Figure A1, which plots grade quantiles for different programs against each other, using the 2019-20 data from NUS-Business. The fact that the data cluster around the 45° line indicates that grades have historically been similar across different master's programs for almost the entire range of the distribution. There are differences, but they tend to be small; MSc grades tend to be slightly higher, and there are some very low EMBA and MBA grades. But in the large, there's no obvious clear and strong argument that there *have* been big differences in grades across programs. This leads one to conclude that there's no real reason to *introduce* such differences in any new policy.

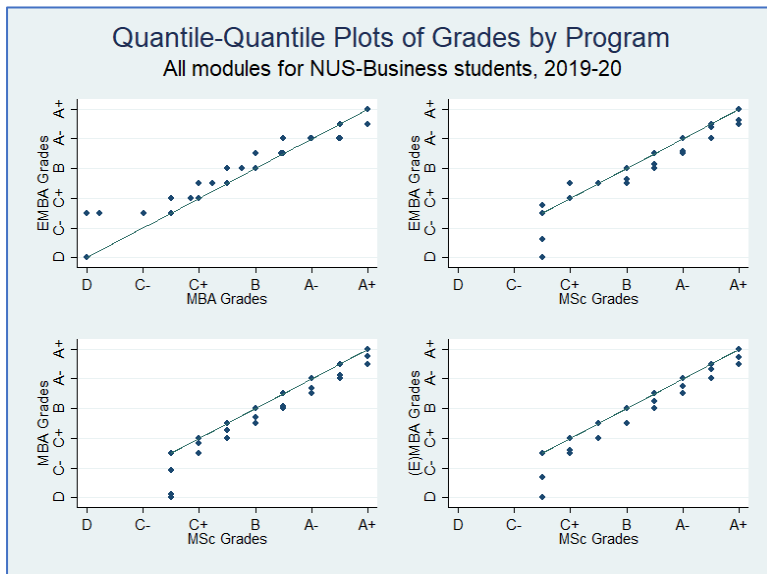


Figure A1

Required and Elective Courses

A different issue is whether there should be fundamentally different schemes for core and elective classes. It turns out, conveniently, that, this dimension does not seem to have had

any dramatic effect on grading outcomes historically. Figure A2 is an analogue to Figure A1, but plots grade quantiles for core (required) courses against quantiles for electives. Figure A2 demonstrates clearly that the nature of a course hasn't made a lot of difference to grade distributions. Again, the data cluster around the 45° line, showing that grades have historically been similar for core and elective courses. Again, this is true for almost the entire range of the distribution, though those low EMBA and MBA grades still show up.

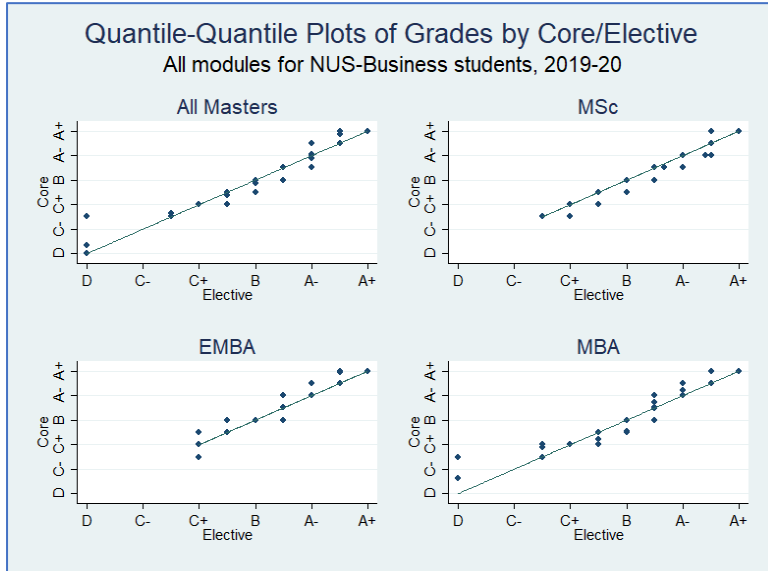


Figure A2

Endnotes

¹ At NUS-Business, these problems are exacerbated since the school's guidelines were widely disregarded, leaving many of students and faculty adrift. In 2019-20, over a third of NUS-Business masters' modules exceeded the school's recommended grade average *maximum*, while three-quarters exceeded the recommended grade *average*. It seems likely that some faculty who stick to the rules are punished for good behavior. It is also relevant that master's grades are not very much used (which makes grade obsession mysterious).

² Obviously, it is too early to have much data from the new coarse system at NUS-Business, and switching the grading system is expected to have consequences for student conduct. That is, the change in policy is expected to affect behavior endogenously, and this is not (yet) fully observable. What I ask is "What would it have looked like if we had assessed students in the past using the fine grading system, but surreptitiously graded them with a coarse system?" What I cannot do is bring any substantive data to bear (yet) on the questions of whether a coarse system has encouraged adventurousness, teamwork, extra-curricular engagement and lower stress without any substantial negative effects on academic motivation or achievement. While this seems plausible, this caveat should be kept in mind.

³ Like many institutions of higher learning, NUS does not have a "D-" grade. For reasons unknown to the author, NUS also does not use "C-".

⁴ The exceptions include courses that are: a) taken on a satisfactory/unsatisfactory basis, and are not assigned a letter grade, often because they are taken overseas; or b) audited. Excluding such cases from the 8,100 student-module observations, reduces the number of graded student-modules to 6,950 observations.

⁵ For reasons unknown to the author, at NUS both A+ and A are worth 5.0 grade points.

⁶ This is an analogue to the more common "Grade Point Average" or "GPA" but was given a different name to emphasize the different scale, since a CAP can maximally be 5.0 but a typical maximum GPA can only be 4.0.

⁷ <https://www.hbs.edu/mba/financial-aid/Pages/satisfactory-academic-progress.aspx> .

⁸ It is reasonable to ask why I examine a three-cell system rather than, say, a still-coarse system with more cells (say, four). There are at least three reasons. First, Ockham's razor encourages simplicity. Second, shifting from a fine to a three-cell system makes for more drama, useful for marketing the innovation and attracting applicants; it is a decisive break from the past in encouraging students to spend time learning and collaborating, not competing and obsessing about grades. Finally, as anyone who's been affected by a bureaucracy (i.e., those sentient) knows, it's always easier to add complication than to simplify.

⁹ The ten NUS grades do not map into three HBS cells perfectly without splitting the former. Still, a rough correspondence is remarkably close, particularly for electives. A more exact mapping would have Pass at 10% and the Distinction capped at 20/25% for core/electives, after taking into account integer constraints. Small fractions of NUS-Business core A and B- grades should be Merit rather than Distinction and Pass respectively; the electives correspond almost exactly to the mapping. Given the endogeneity issue, it seems pointless to do more.

¹⁰ Happily, though not coincidentally, the Distinction, Merit and Pass grading options already exist at NUS on official transcripts. Old (fine) and new (coarse) grade points remain on the same scale, easing implementation and integration with other NUS units.

¹¹ The resemblance across programs is a consistent pattern for all that follows; more is presented in an appendix.

¹² Since 61% of the grades were received for core classes, the cutoff for distinction should be around the 22nd percentile, not the 25th. Similarly, at the lower end, the B- cutoff is at the 11th percentile, not the 10th.

¹³ It's impossible to do meaningful scatterplots for the obvious reasons.

¹⁴ Skewness is $-.41$, a level insignificant at standard confidence levels. There is evidence of leptokurtosis at 4.55 , which is desirable in this context.

¹⁵ We have no reason to believe that selection bias taints the responses.

¹⁶ In contrast to the grading scheme considered here, which mostly compares students relative to each other, Sadler (2005) examines absolute or criteria-based grading.

¹⁷ Intuition is most easily provided by an example with two students who compete with each other on the basis of their relative standing. Suppose student 1 achieves a random mark of between 70 and 80 if she shirks and between 80 and 90 if she works. This is always higher than the result for student 2, who gets between 40 and 50 if he shirks, and between 50 and 60 if he works. With fine grading, neither has an incentive to work, so both shirk. But suppose the professor assigns "A" to scores above 85, "C" to those below 50, and "B" for those in between. In this case, both 1 and 2 are inspired to compete and thus both work. The results of Dubey and Geanakoplos are, of course, far more general than this specific example.