Do Grades Enhance learning? Evidence from a Natural Experiment in a Liberal Arts College

Scott Lee Chua[a], Guillem Riambau[1b], and Tim Wertz[c]

[a]National University of Singapore
[b]Universitat de Barcelona and IEB
[c]Yale-NUS College

June 7, 2021

Abstract

This paper examines how grades affect students' learning using data from a liberal arts college in Singapore. We examine a natural experiment where, in 2017, the grading policy for an introductory statistics module was exogenously changed from letter grade to pass/fail, but the content, format, difficulty, and modes of assessment remained exactly the same. Our dataset includes all results from individual twice-weekly quizzes and two exams for five cohorts — three under letter-grade and two under pass/fail. Our results show that pass/fail students initially work as hard as letter-graded students, but tend to lower their effort as the course goes on — with a sharp decrease right after the first exam. This suggests that, as soon as students realise they have secured a "pass", they choose to allocate their time and effort to other subjects or activities. We find that intending to major closely related to the subject does not explain relatively better performance under a pass/fail scheme. This suggests that students who exert consistent effort regardless of the grading scheme are either those who are naturally gifted or those who are driven by personal ambition (whether to learn, score well, or outperform their peers). Our results suggest that pass/fail policies could safely be implemented more extensively: the loss in acquired knowledge is marginal and could easily be compensated in terms of personal time and mental health.

Keywords: Education; Grades; Effort; Learning.
JEL Codes: I21, I23
Authors' declarations of interest: None.
Funding: We gratefully acknowledge funding from Yale-NUS College.

[1]Corresponding author. Email: griambau@gmail.com.

1 Introduction

Many colleges throughout the world implemented a pass/fail policy for their courses under the COVID-19 pandemic in 2020 (Basken, 2020). This means that, instead of getting a traditional letter grade (A, B,..., F) students either pass the course or fail it, and such is the only information kept in their transcripts. This policy is, however, not new: the Massachusetts Institute of Technology evaluates all first-year students on a pass/fail basis (MIT Registrar's Office, n.d.), and medical schools around the world have implemented pass/fail to various degrees (White & Fantone, 2010).

Beyond the extraordinary COVID-19 scenario, the usual rationale for implementing such policy is to encourage a learning-oriented attitude in students (Dahlgren et al., 2009) with the goal of nurturing lifelong learners (Jacobs et al., 2014). Another reason for implementing a pass/fail policy has to do with students' mental health: Previous studies have shown that stress levels are reduced (Spring et al., 2011) whereas satisfaction with the program increases (Bloodgood et al., 2009).

Less is known about how a pass/fail policy affects actual learning outcomes. Do students learn as much when the stakes are lower, as it is the case under a pass/fail system? We take advantage of a change in grading policy that took place at Yale-NUS College, a liberal arts college in Singapore, to examine this question in detail.

All incoming students at Yale-NUS College have to take a mandatory statistics module called "Quantitative Reasoning" in their first year. Until 2017, grading was based on the standard letter system. From academic year 2017-18 onward, this module became pass/fail. Importantly, (i) this decision was completely exogenous to students and professors involved in delivering it; (ii) all learning outcomes, teaching materials, and assessment tools remained virtually the same; (iii) first-year students are very similar across years. This allows us to compare learning across cohorts and examine whether a change in grading policy affects learning outcomes.

Our natural experiment has two further advantages. First, students take twenty individual quizzes (two per week) that count toward their final grades, allowing us to observe students' performance *over the course of a semester* in high fidelity. Second, students' final (numerical) grades are still computed even under the pass/fail policy, providing us with clean counterfactuals.

Although students are never officially told the numerical grade cutoff that secures a "pass," they are aware that only a handful of students (around 5–10%) fail each year. In other words, students are aware that, unless they are substantially below average, they will pass. In particular for this module, students have ample opportunity to assess how well they are doing relative to their peers. For instance, when multiple choice quizzes are conducted in class, students receive instant feedback on the proportion of their classmates that answered each question correctly. Thus, we can reasonably assume that students can (and do) assess their relative performance with their peers' throughout the semester.

We have data on five cohorts, each comprised of 200 students on average. The first three cohorts took Quantitative Reasoning on a letter-grade basis, whereas the last two took it on a pass/fail policy. In our analysis, we can control for gender, intended major at the point of admission to the college, and country of origin.

Our results show that the grading scheme unambiguously matters: letter-graded students get higher final scores on average than their pass/fail peers. However, quiz and exam scores in the first half of the semester are on average virtually identical across all cohorts. Pass/fail students only exhibit lower scores on quizzes and exams in the second half of the semester. We interpret this as follows. Students exert a high effort at the beginning of the course because they are uncertain about what a "pass" entails: they are new to an international college environment and have little to no prior information on the difficulty of a module. However, as the semester progresses, when students learn that they are comfortably above bar, they re-allocate their time and effort into other subjects or activities. As a result, scores for quizzes taken at the second half of the course, as well as for the second exam, are substantially lower for the pass/fail cohorts.

This behavior is by no means universal. We observe that a handful of students still get (nearly) maximum scores throughout the whole semester under the pass/fail system. There are four potential reasons that could explain the behavior of such students: (i) they instrumentally exert a higher effort since material in this module will benefit them later, as would be the case for prospective economics or mathematics students, for instance; (ii) they genuinely enjoy the topics taught in the module, which —given that we do not directly observe students preferences— we again proxy by intended major; (iii) they are naturally talented and easily master the materials regardless of effort; and (iv) they are intrinsically motivated to get the highest possible scores, either to learn, feel accomplished, or to outperform their peers.

We do not find any evidence that (i) or (ii) could be driving the behavior of students who get (nearly) maximum scores throughout the whole semester. All our specifications show that intended field of study is uncorrelated with performing exceptionally well under pass/fail. This suggests that one (or both) of mechanisms (iii) or (iv) must be at play: either natural talent or personal ambition explain why some students get remarkable scores regardless of the grading scheme.[1]

The implications of our findings are remarkable. Under pass fail, we do not find bunching right above the "pass" mark. The distribution of grades under pass/fail is quite similar to that under letter-grading —it just so happens that the mean is relatively lower, so that the distribution is mildly shifted to the left. Thus, the extent to which letter grades improves academic performance should also not be exaggerated.

This allows us to the address the following normative, policy-relevant question: *How desirable is a pass/fail scheme?* Clearly, there are learning losses under a pass/fail. However, these are marginal, and could easily be compensated by gains in terms of mental health, stress, and re-allocation of time and effort by students. Overall, in our view, results presented here suggest that universities could safely embark into pass/fail policies for multiple courses with minimal consequences upon their students' academic learning.

2 Context & hypotheses

2.1 Institutional context: Yale-NUS College

Yale-NUS College (located in Singapore) was inaugurated in academic year 2013-14, after a few years of preparatory collaborative work between the National University of Singapore (NUS) and Yale University. While it is an independent liberal arts institution, with its own governing body and president, it is located within the wider NUS campus and shares resources and policies with NUS. Yale-NUS College administers a four-year program, where each academic year is divided into two semesters, and four modules (i.e., classes) are considered a full semestral load. All students are required to take a set of ten introductory modules called the "Common Curriculum," seven of which must be taken in the first year. That is, all first year students take the same four modules in their first semester, and the same three modules (plus an optional elective) in their second semester.

All four modules of the first semester of year 1 operate under a pass/fail basis. From semester 2 onward, all courses follow a standard letter-grade grading scheme (A – B – ... – F).

[1]Unfortunately, the college holds no records that would allow us to falsify either (iii) or (iv).

The official rationale for this is to lower the stakes for students, and allow a smoother transition into university.[2]

One of the compulsory modules is "Quantitative Reasoning," a subject that "addresses what might be called numeracy or quantitative literacy by exploring some basic topics related to algorithmic thinking and statistical inference" (Yale-NUS College, n.d.). This module is taught by a team of around ten professors, where each professor handles a section of around 25 students. Subject to criteria on gender, region, and stated intended major, students are randomly assigned to sections. All readings, classroom materials, assignments, exams, and rubrics are exactly the same across sections. Professors meet their section twice a week for thirteen weeks, and have no room to deviate from the syllabus, as all sections are expected to cover the same material and quizzes in a given week. The syllabus from academic year 2016-17 can be found in Appendix 1 of the supplementary materials.

2.2 Change in grading policy

When the college began operations, Quantitative Reasoning was held in the second semester of the first year (i.e., students would be letter-graded). In 2017, an external committee conducted a pedagogical review of the curriculum — a review that had been scheduled since the college's inception. The committee suggested that Quantitative Reasoning be moved to the first semester of the first year (i.e., students would be graded pass/fail). Table 1 summarizes the grading policy students in each cohort faced over the years.

Table 1: Grading policy by cohort.

Academic year	Grading Policy	Semester
2014-15	Letter grade	2
2015-16	Letter grade	2
2016-17	Letter grade	2
2017-18	Pass/Fail	1
2018-19	Pass/Fail	1

Two features of this natural experiment are key to our identification strategy. First, the change in grading policy was exogenous — both in the sense that incoming/prospective students had no say in it, and in the sense that the course itself (syllabus, rubrics, exams, course materials)

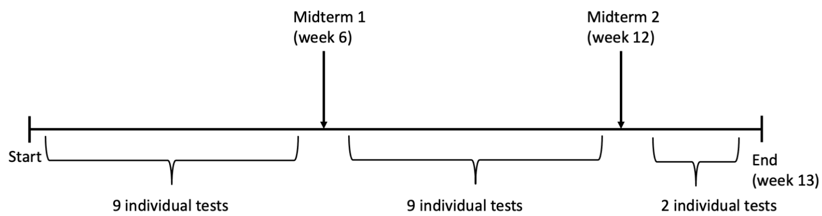
[2] Anecdotally, however, some claim that the real rationale behind this policy —which is similar to the one at NUS — is to allow male students to readjust to academic habits after a two-year long military service. In Singapore, all males *without exception* have to fulfill the military service when they turn 18. This means that Singaporean women enter college right after high school at the age of 18, while men enter college when they are 20, after a two-year gap spent in the army.

remained unchanged. Second, students' numeric (raw) scores are always computed, regardless of the grading policy. This ensures a clean counterfactual. The change in grading policy mainly affected what was *recorded* on students' transcripts — a letter grade, or a pass/fail mark. Students always knew in advance which grading policy they faced, and could not opt-out (e.g., a student who scored an A under the pass/fail policy could not convert their "Pass" into an "A" on their transcript).

2.3 Details of the module

Quantitative Reasoning is a one semester course that follows a flipped-classroom model, meaning that students are expected to read most materials before class, and apply them during individual or group activities during class. At the start of each class, students take an individual multiple-choice quiz based on the assigned readings. As these quizzes are administered via an online portal, they are automatically timed (for around ten minutes) and provide students with immediate feedback. There are twenty such quizzes over the course of a semester. After the lowest two quiz scores are discarded, the remaining eighteen are equally weighted to comprise 10% of a student's final grade. Further details on class dynamics and in-class exams can be found in Appendix 1.

Figure 1: Detailed timeline of Quantitative Reasoning



Students also take two exams, in week 6 and week 12 (see Figure 1 for a detailed timeline). Each exam consists of two parts: 20 multiple-choice questions (60 minutes) and a coding exercise in **R** (30 minutes). As a rule, students are informed of their exam scores in the subsequent class. Since teaching materials have been kept consistent through the years, exams have barely changed. Overall, 80% of the questions remain identical from one year to the next, 10% are reworded or renumbered but remain essentially the same, and the remaining 10% of questions are created from scratch each year. Reasons for rewording or substituting exam questions vary. Often, questions which are considered too imprecise, ambiguous, hard, unpolished, or unrelated to course materials are changed. This would generally make exams easier over time. Other times, however, questions are dropped because they are considered too easy, or simply as a matter of taste (i.e., as the teaching team changes over time, new professors may prefer to ask different questions). We are confident that, despite minor changes, exam difficulty has remained stable over time.

While exams barely change from one year to the next, exam questions cannot possibly be shared with incoming cohorts. The multiple choice questions part of the exam is always done on paper, and are graded by hand. Students have the opportunity to review their exam after it has been graded, but they are not allowed to keep, photograph, or even remove it from the professor's office. Furthermore, breaching the academic honor code at Yale-NUS College carries severe sanctions of which students are fully aware. For these reasons we are confident that there is no "black market" for exam questions in the college. If that were the case, we would find performance to keep improving over time, which we do not (even when we look only at the letter-graded cohorts, or only the pass/fail cohorts).

On the other hand, students can keep the questionnaire for the coding exercise, as this portion of the exam is completed on their personal laptops. This matters little as students are told in advance (and in detail) the kind of tasks they will be asked to perform (e.g., compute the mean, draw an histogram, perform a t-test, etc.). While questions are in essence the same from one year to the next, the particular dataset students are provided with changes every year.

Beyond quizzes and exams, there are other assessments that count toward the final grade (e.g., a group presentation, a short paper, class participation). In this paper, we only use the grades for quizzes and exams, since they are the only ones that can be directly and objectively compared. More details on the grading scheme of the course can be found in Appendix 1.

2.4 Expectations

Given the change in policy summarized in Table 1, we expect students to perform relatively worse when Quantitative Reasoning is graded pass/fail (i.e., *not* letter-graded). However, we expect this to be the case only for the latter half of the course. For the first half of the course, we expect performance to be roughly equal regardless of the grading policy.

The reason for this is as follows. At the start of the semester[3], students are informed that only around 5–10% of students each year fail the class (under either grading policy). Hence, for those cohorts under the pass/fail grading policy, students know that, as long as they do not

[3] Beginning with the second cohort.

perform dramatically below average, they will pass the course. However, first-year students have no previous experience in college. As an international liberal arts college, Yale-NUS students come from very distinct cultural and academic backgrounds. This means that, even if students know they only need to make sure they are not too far behind other students, they do not really know what that entails. In other words, the amount of effort they need to exert in order to secure a "Pass" is unknown to them.

As weeks go by, they get a sense of how well they are doing, and so they learn what is expected from them to pass. In Quantitative Reasoning in particular, students receive feedback in the form of a quiz score twice a week, which they can easily compare with peers. Within weeks, they get a sense of whether they are doing above, around, or below average. After the first exam, which is their first sizeable assignment (15% of the final grade), they have a very accurate idea of where they stand.

As a result, we expect all cohorts to exert a high effort at the beginning of the course, when requirements are still unknown. However, we expect different behaviors after the first sizeable grading signal (in this case, the first exam): we expect students in the letter-graded cohorts to keep exerting a high effort after the first exam, and students in the pass/fail cohorts to lower their effort after the first exam. For the pass/fail cohorts, getting an A, B, or C all result in a "Pass" on their transcripts, so we expect students to reallocate their time and effort toward other subjects or activities they find more rewarding or enjoyable.

Of course, we do not expect all students to behave in a same manner. Some may still exert a consistently high effort all the way through, even under a pass/fail policy. There are two types of students who may behave in such a manner: (i) students whose intrinsic motivation is higher than average (be it because they want to learn as much as possible, or because they want to remind themselves they are top students in their cohort); and (ii) students who are highly motivated for the particular subject (be it because they can instrumentally benefit later during their intended major, or because they genuinely enjoy it).

Finally, there are some talented students who do exceptionally well in the course even if they do not exert a high effort. We expect these students to perform extraordinarily well regardless of the grading scheme.

We can summarize our expectations in the form of three hypotheses:

- *Testable Hypothesis 1: Effort will be similar across cohorts until the first exam. Effort will be lower for pass/fail cohorts after the first exam.* We proxy effort with test and exam scores: in a Difference-in-Difference approach, we expect scores for the pass/fail cohorts to be comparatively lower after the first exam.

- *Testable Hypothesis 2: For students who do exceptionally well until the first exam, we expect scores to remain high throughout the course regardless of the grading policy.* "Exceptionally well" is defined as students who are in the top five percentile of their cohort. These are students who are particularly talented for the subject, or whose intrinsic motivation is remarkably higher.

- *Testable Hypothesis 3: Students with a high motivation for the subject will exert a high effort throughout regardless of the grading policy.* In our case, since we do not have detailed information on student preferences, we proxy motivation for the subject with a dummy for whether a student intends to major in economics, physical sciences, or mathematical, computational, and statistical sciences (MCS) at the point of admission to the college.

2.5 Data overview

All data we use is official administrative data from Yale-NUS College. We have data for 3 cohorts that took Quantitative Reasoning under a letter-grade policy (2014-15/2015-16/2016-2017), and for 2 cohorts that took if under a pass/fail policy (2017-18/2018-19).[4] For anonymity purposes, our data includes information only on cohort, gender, nationality,[5] and intended major of study at the time of the entrance to the college. Table 2 summarizes student demographics.

On the other hand, our data on test and exams scores is much more detailed. For each student, we have the results for all individual quizzes done in each class, as well as for both exams. We also have data for all other scores earned throughout the course (see Appendix 1 in the supplementary materials for more details), but we do not use those as they cannot be systematically compared across years. Table 3 summarizes student scores by cohort and grading policy.

[4]We do not use the first cohort because there are substantial changes between the first year and the following ones.

[5]For anonymity reasons we have to group this into regions: Singapore ($N = 603$), East Asia($N = 94$), Europe ($N = 53$), North America, ($N = 73$), South Asia ($N = 78$), Southeast Asia ($N = 64$), and Other ($N = 67$).

Table 2: Student demographics by grading scheme.

	Cohorts	
	Letter-grade	Pass-Fail
Female	56.3%	54.9%
Region		
Singapore	62.2%	56.7%
East Asia	9%	9.6%
Southeast Asia	6.6%	6%
South Asia	5.3%	8.3%
Europe	3.8%	6.9%
North America	7.9%	6.4%
Other	5.3%	6%
Intended major		
Sciences	18.3%	19.1%
Social sciences	54.1%	47.6%
Humanities	27.3%	29.5%
Undecided	0.4%	3.7%
Number of students	531	481

Sciences: Life Sciences, Mathematical, Computational and Statistical Sciences (MCS), Physical Sciences. Social sciences: Anthropology, Double Degree Programme in Law and Liberal Arts, Economics, Environmental Studies, Global Affairs, History, Psychology, Urban Studies. Humanities: Arts and Humanities, Literature, Philosophy, Politics, and Economics (PPE), Philosophy. Letter-grade cohorts: 2014-15; 2015-16; 2016-17. Pass-fail cohorts: 2017-18; 2018-19.

Table 3: Average scores by cohort and grading policy.

	Cohorts	
	Letter-grade	Pass-Fail
Quizzes *before* exam 1	60.9	68.5
Quizzes *after* exam 1	66.8	66.6
Exam 1	79.6	76.4
Exam 2	80.3	67.9

9 quizzes before the first exam. 13 quizzes after the first exam. Letter-grade cohorts: 2014-15; 2015-16; 2016-17. Pass-fail cohorts: 2017-18; 2018-19.

3 Results

In order to test our main hypothesis, we use the following main specification:

$$y_i = \alpha + \beta_1 Lettergrade_i + \beta_2 Post_i + \beta_3 Lettergrade_i \times Post_i + \gamma X_i + \varepsilon_i, \quad (1)$$

where subscript i refers to an individual student. y_i is student i's exam or average quiz score; *Lettergrade* is a dummy indicating that a student belongs to a letter-graded cohort; *Post* refers to scores strictly after the first exam; and X_i is a vector of student-specific controls (gender,

Table 4: Regression results: exam scores

	All obs.	Women	Men
Letter grade	0.597	-2.624*	5.757***
	(1.123)	(1.487)	(1.547)
Exam 2	-8.601***	-9.305***	-7.685***
	(0.814)	(1.127)	(1.161)
Letter grade ×	9.260***	11.035***	6.921***
Exam 2	(1.723)	(1.689)	(1.952)
Observations	2,008	1,117	891
R^2	0.266	0.172	0.213
Controls	YES	YES	YES

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors (clustered at the cohort level) in parentheses. Controls: gender, intended major, region, cohort.

major, region, cohort).

Recall that our hypothesis is that effort (proxied with quiz and exams scores) will be similar across cohorts until the first exam, while effort will be lower for pass/fail cohorts after the first exam. Hence, β_3 is our coefficient of interest: If our hypothesis is correct, we should find that β_3 is positive and significant. Results are displayed in Tables 4 (for exam scores) and 5 (for average quiz scores). Both tables unambiguously show that effort seems to fall for pass/fail cohorts after the first exam. In Table 4, this is shown by a positive and significant coefficient on the interaction term "Letter grade × Exam 2." On average, letter-graded students score 9 points higher (out of 100) on the second exam than their pass/fail peers. The effect holds for both genders, although it seems to be stronger for female students: those in the pass/fail cohorts tend to score 11 points less than those in the letter-graded cohorts (for male students, 6 points less).

Table 5 shows the results when we use quiz scores. The relevant coefficient is the one associated to "Letter grade × Post Exam 1." On average, letter-graded students score 8 points higher (out of 100) on quizzes after exam 1 than their pass/fail peers. The effect holds equally for both genders, with minimal difference in effect size for women versus men.

In order to test our second hypothesis, we run the following triple-differenced specification:

Table 5: Regression results: average quiz scores

	All obs.	Women	Men
Letter grade	0.758	-0.521	3.721**
	(1.007)	(1.348)	(1.496)
Post Exam 1	-1.921**	-1.629	-2.277**
	(0.757)	(1.024)	(1.134)
Letter grade ×	7.909***	8.029***	7.734**
Post Exam 1	(2.051)	(1.641)	(2.731)
Observations	2,024	1,126	898
R^2	0.321	0.321	0.256
Controls	YES	YES	YES

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors (clustered at the cohort level) in parentheses. Controls: gender, intended major, region, cohort.

$$y_i = \alpha + \beta_1 Lettergrade_i + \beta_2 Post_i + \beta_3 Top5\%_i + \beta_4 Lettergrade_i \times Post_i +$$

$$\beta_5 Lettergrade_i \times Top5\%_i + \beta_6 Post_i \times Top5\%_i + \beta_7 Lettergrade_i \times Post_t \times Top5\%_i + \gamma X_i + \varepsilon_i, \quad (2)$$

where "Top 5%" indicates that a student did better than 95% of their cohort in the quizzes leading up to exam 1, and y_{it} refers to either student i's first ($t = 1$) or second ($t = 2$) exam score. For the sake of simplification, we present the results from the predicted scores (using the "margins" command in Stata) in Table 6 below.

Table 6: Effect of grading policy change on top students' grades.

	Δ grade from exam 1 to exam 2	Difference
Top students, letter-graded	-11	-3
Top students, pass/fail	-14	
All other students, letter-graded	+2	-11
All other students, pass/fail	-9	

Δ grade means change in predicted grade: "-11" means that, for "top" students, the average predicted grade for the second exam under the letter-grade policy was 11 points lower than for the first exam.

We can see that top students' scores were barely affected by the policy: top pass/fail students lose only three points more on average in the second exam than the first exam, compared to top letter-graded students. However, the effect is much larger for all other students. All other letter-graded students saw a marginal improvement in score from the first to second exam, while pass/fail students do 9 points worse. Results with quizzes scores (not shown) reveal a similar pattern.

Finally, we check the behavior of students whose intrinsic motivation may be higher than average because of their preference for the subject. We run the same specification as in (2), where instead of "Top 5%" we include a dummy for having indicated that their intended major is economics, physical sciences or MCS (results not shown). We find no effects whatsoever. If we test these intended majors separately, results do not change. This suggests that students preferences over their major plays little role in intrinsically motivating them, even when the subject at hand is substantially more related to some majors than to others.

4 Concluding remarks

This paper has presented evidence that shows how students' learning is modified under different grading schemes. Higher stakes grades[6] (such as letter grades) result in marginally better performance than lower stakes schemes (such as pass/fail). This suggests that learning is indeed higher under higher stakes schemes.

Nonetheless, results in this paper far from suggest that letter-grading is a superior grading scheme. As discussed above, pass/fail has been show to improve satisfaction and mental health among students. Furthermore, results presented here only refer to short-term learning. We have no basis to show that those marginal gains in the short run are sustained in the longer run. Further research focused on the longer run effects should shed more light on this policy-relevant issue.

[6]Hiring and base salary for junior positions in Singapore is highly contingent on GPA university scores. In this sense, grades for college students in Singapore are indeed high stakes.

References

Basken, P. (2020). *Us colleges adopt pass-fail rules, stirring wider reform*. Retrieved May 30, 2021, from https://www.timeshighereducation.com/news/us-colleges-adopt-pass-fail-rules-stirring-wider-reform

Bloodgood, R., Short, J., Jackson, J., & Martindale, J. (2009). A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Academic Medicine*, *84*(5), 655–662.

Dahlgren, L., Fejes, A., Abrandt-Dahlgren, M., & Trowald, N. (2009). Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education*, *14*(2), 185–94.

Jacobs, J., Samarasekera, D., Shen, L., Rajendran, K., & Hooi, S. (2014). Encouraging an environment to nurture lifelong learning: An asian experience. *Medical Teacher*, *36*(2), 164–68.

MIT Registrar's Office. (n.d.). *Grading policies*. Retrieved May 30, 2021, from https://registrar.mit.edu/classes-grades-evaluations/grades/grading-policies

Spring, L., Robillard, D., Gehlbach, L., & Simas, T. (2011). Impact of pass/fail grading on medical students' well-being and academic outcomes. *Medical Education*, *45*, 867–77.

White, C., & Fantone, J. (2010). Pass-fail grading: Laying the foundation for self-regulated learning. *Advances in Health Sciences Education*, *15*, 469–77.

Yale-NUS College. (n.d.). *Common Curriculum: Quantitative Reasoning*. Retrieved May 30, 2021, from https://www.yale-nus.edu.sg/curriculum/common-curriculum/quantitative-reasoning/