EDUCATION

AMAZING GRADES

Grade Inflation at New Zealand Universities

James Kierstead
with Michael Johnston
Foreword by Prof Douglas Elliffe

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

THE NEW ZEALAND INITIATIVE

Published August 2025 by
The New Zealand Initiative
PO Box 10147
Wellington 6143
New Zealand
www.nzinitiative.org.nz

ISSN
2624-0092 (print)
2624-0106 (online)

RR92

Printing arranged by TBD Digital

AMAZING GRADES

Grade Inflation at New Zealand Universities

James Kierstead
with Michael Johnston
Foreword by Prof Douglas Elliffe

About The New Zealand Initiative

The New Zealand Initiative is an independent public policy think tank supported by chief executives of New Zealand businesses. We believe in evidence-based policy and are committed to developing policies that work for all New Zealanders.

Our mission is to help build a better, stronger New Zealand. We are taking the initiative to promote a prosperous, free and fair society with a competitive, open and dynamic economy. We are developing and contributing bold ideas that will have a profound, positive and long-term impact.

ABOUT THE AUTHORS

James Kierstead is a Research Fellow at the New Zealand Initiative, where his work focusses on universities, free speech, and democracy. Born in Canada, he earned a BA in classics at Oxford and a Master's in ancient history in London before moving to Stanford, where he earned another MA in political science and completed his PhD in classics in 2013. That year he also moved to Wellington to take up a position as a Lecturer (later Senior Lecturer) in classics at Victoria University. In 2022 he began working at the New Zealand Initiative, publishing his first report *Blessing or Bloat? Non-academic Staffing at New Zealand Universities in Comparative Perspective* (co-authored with Michael Johnston) in August last year, a few months after which his role at Victoria was disestablished. Last year he published his second report, *Unpopular Opinions: Academic Freedom in New Zealand*, which helped make the case for the academic freedom legislation currently going through Parliament. Besides his academic publications on ancient and modern democracy, James has published pieces in *The Post, The New Zealand Herald, The Australian, The Spectator Australia, Quillette, Quadrant, Times Higher Education, Chronicle of Higher Education* and other outlets. He is also the co-host, with Michael Johnston, of the Free Kiwis! podcast, which aims to establish liberalism in New Zealand.

Dr Michael Johnston is a Senior Fellow at The New Zealand Initiative. He leads the Initiative's work on education. He is a cognitive psychologist with a background in literacy research, educational assessment and psychometrics. Prior to his time at the Initiative, he was the Associate Dean (Academic) of the Faculty of Education at Victoria University of Wellington. Between 2005 and 2011, he worked at the New Zealand Qualifications Authority (NZQA), where he developed a new, more reliable marking system for NCEA examinations. In 2024, Michael chaired a Ministerial Advisory Group (MAG) for Education Minister Erica Stanford. The MAG advised on the development of a knowledge-rich curriculum for English and mathematics. Following that work, Michael is currently a member of the Curriculum Coherence group, which advises on the development of knowledge-rich curricula across all school subjects. Michael is also a member of a technical advisory group to NZQA on assessment for NCEA and New Zealand Scholarship. In his time at The New Zealand Initiative, Michael has published reports on Modern Learning Environments, systems reform in education, teacher education, the use of AI in education, and pathways for industry training and apprenticeships.

Contents

Figures

Tables

PREFATORY NOTE

This report was written primarily by me, James Kierstead, with substantial support and input from my colleague Michael Johnston (see further below). I realised at a late stage in the report that I had switched several times between writing in the first person singular ('I') and the first person plural ('we'). To avoid confusion, 'I' refers to James and 'we' tends to refer to James and Michael.

I have pluralised letter grades (A, B, C etc.) as 'As, Bs, Cs' and so on except at the beginning of sentences, where 'As' might be mistaken for the word 'as' with a capital 'A.' In those cases I have written 'A's' with an apostrophe between the two letters.

ACKNOWLEDGEMENTS

I am especially grateful to Michael Johnston for helping analyse the data and for sharing his expertise in psychometrics and his experience from his days at the New Zealand Qualifications Authority (NZQA). Darwyyn Deyo and Jemma Stevenson gave me some generous and helpful advice about data analysis. Darwyyn also helped me think through 'the grade inflation game' (see Chapter 4), as did Tony Burton. Benno Blaschke helped me a number of times with my prose. I am grateful to the rest of the Initiative team for their comments at a 'brown bag' session on grade inflation, and specifically to Oliver Hartwich, Chelsy Killick and Jamuel Enriquez for helping me frame, develop and (eventually) complete this report. Finally, I thank the staff members at each of New Zealand's universities who provided data on grade distributions and also responded to follow-up queries and requests for more or different data.

Foreword

'Grade inflation' occurs when higher grades are awarded for the same level of academic achievement. It has been the subject of concerned discussion in universities for at least as long as I've worked in one. Are we passing students' work that doesn't deserve to pass? Are we conferring distinctions on good but not exceptional work, thus implying a level of achievement that hasn't actually been demonstrated? If we are doing that, are we failing to give the best work the recognition it deserves?

I've in the past been sceptical about the magnitude of the problem, and even whether it really existed. It's always tempting to look back at the past through rose-tinted spectacles. And it always seemed to me just a little too easy to assume that university (among other things) was harder back then, and that kids nowadays have it easy.

In this report, Dr Kierstead has convinced me that grade inflation is no lazy trope. It is real, shows no sign of abating, and has real costs. Kierstead ably lays out the evidence for shifting grade distributions over time, both in overseas universities and in New Zealand.

He also looks at a number of possible reasons why this shift might have occurred without standards being relaxed – in which case we would have, not grade inflation, but justifiable grade rises. But none of these alternative explanations seems able to account for the grade rises that have taken place at our universities.

In the end, we can't escape the conclusion that an A grade of today doesn't represent the same level of achievement that it did in the past. Kierstead discusses possible reasons that this has come to pass, as well as possible interventions that might reverse, or at least halt, the trend.

One possibility that I think might be worth looking into in more depth is that we've got better at teaching over the years. I certainly like to think that I've got better at recognising and explaining material that students tend to find challenging. And universities definitely take teaching performance more seriously than in the past.

Better teaching probably won't explain all the grade rises laid out in this report (especially the surge in A grades during COVID). What if it does explain some of the increase, though? How worried should I be that my students are getting better results if I've also got better at teaching them?

The answer to that depends on how we look at academic achievement. Is it simply a matter of students acquiring a certain body of knowledge and attaining a certain pre-ordained standard? If so, then the more students are meeting that standard, the more As we should be awarding.

The problem with this more traditional view is that it assumes that educational standards should be static. But perhaps we should be more forward-looking. As we get better at teaching, perhaps we should expect more of our students.

If that's right, perhaps the pattern of rising grades is an opportunity to push towards ever-higher standards. We academics often say that we want our students to outstrip our own achievements. This may be a chance to show that we mean it.

All of this, though, will have to wait until we can determine the extent to which grade rises really are down to improved teaching.

One thing that I think Kierstead is definitely right about is that an ethic of kindness has been an important driver of grade inflation. The impact of COVID on universities and their students probably accelerated this, even if, as this report makes clear, the spike in top grades during COVID is only one part of a larger pattern of grade rises.

But COVID brought to the fore a feeling that we as academics should be more sensitive to the various difficulties that our students face, give more attention to pastoral care, and avoid damaging our students' sense of their own worth. It's not surprising that we may have reacted to that with the apparently simple solution of grading more generously.

When we say 'what a lovely picture, darling' to our toddler's latest unrecognizable splodge, we're being kind, and that may well be the right approach when we're dealing with toddlers. But does that make it right to say, in effect, 'what a splendid essay' to a young adult when we know that it's not really splendid at all?

It may seem kind, but it denies them the pleasure that comes from knowing that the praise they're receiving is based on real achievement. Lack of honest feedback, it seems to me, is only superficially kind. True kindness involves challenging our young people to reach higher levels of achievement. And the satisfaction produced by real achievement isn't something that unconditional praise can provide.

Kierstead closes his report with some suggestions of 'easy wins' – simple policies that universities could adopt right now to start mitigating the worst excesses of grade inflation. So let me close this foreword by suggesting an easy win of my own.

I find it disquieting to watch graduation ceremonies at which more than half the students receiving a given degree are awarded first-class honours. Receiving a first should be special, and a mark of having achieved at a level that was beyond most of one's classmates.

We could correct this right now by simply adjusting the definition of a first to A and A+ averages only, rather than including A- in this as we do now. Why don't we?

In sum, this report gave me plenty of food for thought – and for some uncomfortable reflection. It provides evidence that the problem of grade inflation is worse than I thought it was.

And it made me wonder whether, in the name of kindness, we are denying those who have achieved something genuinely special the right to feel genuinely special. If failure is impossible, success becomes meaningless. We need to restore to our students the right to feel pride in true achievement, and we can only do that by evaluating their accomplishments honestly.

Prof Douglas Elliffe
Professor of Psychology
University of Auckland

CHAPTER 1

Grade Inflation and its Discontents

Grade inflation: a personal tale

It was the end of my first term as a lecturer, and I was exhausted. I had just finished teaching a big first-year lecture course on ancient Greek civilization and a smaller Latin class in which we worked through sections of Lucretius' notoriously difficult first-century BC poem *On the Nature of Things*.

Now it was time for the grading. That was going to be difficult enough, with nearly two hundred exam scripts stacked on my desk. As I began wading through them, though, I was given an additional challenge. I was told that I must not, under any circumstances, fail more than 20% of registered students. Otherwise the government would withdraw funding for the course.

The difficulties didn't end there. In the first meeting of the following trimester, I was told, in no uncertain terms, that my grading for the Latin course had been too severe.

I quickly began picking up – and adapting to – the grading culture in my corner of Victoria University. Colleagues took pride in grading Honours students 'holistically.' Grades, I was told, shouldn't be seen as a 'Platonic form.' I was regularly encouraged to put students whose final grade had come within 1%, or even 1.5%, of a grading boundary over that boundary and to award them a higher grade.

Am I accusing my former colleagues of corruption or cynicism? Certainly not. My impression was, and remains, that my colleagues in the Classics Programme were, on the whole, dedicated professionals with a keen interest in the success of the programme and of the students that passed through it.

They also often set standards that were a great deal higher – it always seemed to me – than are expected of most undergraduate students in New Zealand. This was particularly the case with our Latin and Greek students. If I was occasionally encouraged not to be stingy with A grades in these classes, this was because of a perception that these courses were significantly more challenging than most undergraduate courses these days – a perception that I shared.

In any case, if there was corruption (surely too strong a term for generous grading), then it was a corruption that I came to participate in. Why did I not hold firm, applying my own strict grading standards to my students come what may?

I didn't 'hold firm' because holding firm would have been foolish. It would have made me deeply unpopular with my colleagues, and with good reason. In a faculty (Humanities and Social Sciences) that was constantly haemorrhaging students, getting a reputation as a harsh grader might have put our course offerings at risk. The teaching of Latin and Greek is currently being phased out at Victoria because these languages attract (in the university's judgment) too few students, and that would have almost certainly happened years earlier had I 'held firm.' Cuts are also often justified with reference to student numbers at the programme level, which means that taking a 'heroic' stand on grading can put your colleagues at greater risk of losing their jobs.

The way I very quickly came to grade students at Victoria, in other words, was mainly a product of incentives, incentives that were generated by the system I was operating in. This was undoubtedly also the reason that my colleagues had the grading

norms I observed when I arrived at Victoria (along, perhaps, with an ethic of 'kindness').

Though I eventually picked up and followed the grading norms I saw around me during my time at Victoria, questions always remained at the back of my mind. What were the results of these norms on the whole set of grades awarded at Victoria? Was there grade inflation, and if so, how substantial was it? And how did it compare with grade inflation in other similar university systems?

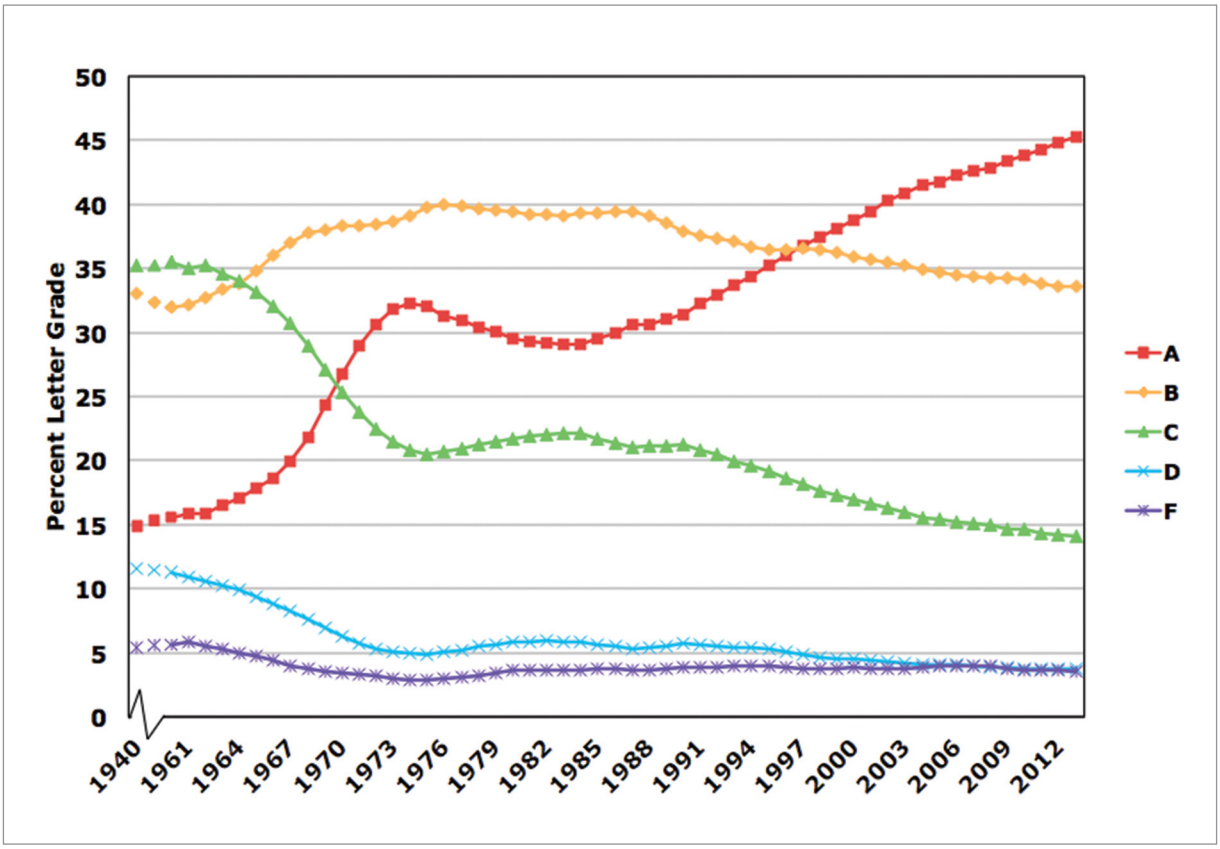
Grade-inflation in other English-speaking university systems

Numerous studies have shed light on grade inflation at universities in other English-speaking countries.

In the US, grades increased by an average of 0.14 Grade Point Average (GPA) points per decade between 1963 and 2013,[1] by 0.28 in total between 1988 and 2013,[2] and by 0.34 between 1990 and 2020. That last figure represents a 12% increase in average GPA, but the median GPA grew even faster, by 21.5%, over the same period.[3]

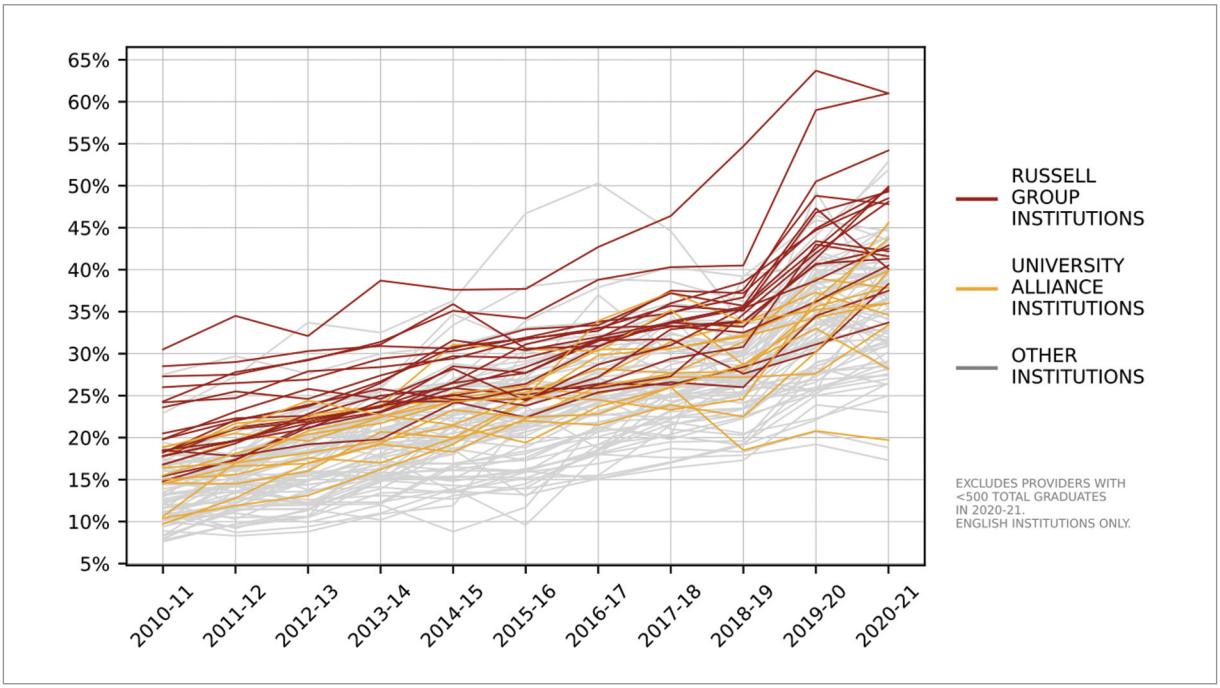
As Figure 1.1 shows, before the Vietnam War, the most common grade awarded at US colleges was a C. The most common grade became a B during that war, and an A shortly before the turn of the millennium. By 2012, the proportion of grades that were in the A-range was 45% – and rising.[4] At Harvard and Yale, just under 80% of all grades are in the A-range as of 2023.[5]

Figure 1.1: Percentage of grades in ranges A to F at four-year colleges in the US, 1940–2012



Source: Stuart Rojstaczer and Chris Healy, "Grade inflation," *op. cit.*

Figure 1.2: Percentage of first-class degrees awarded at universities in England, 2010–2021



Source: Samuel Jones, "New data available on grade inflation," *HE Insights* blog (15 May 2022).

In England, the percentage of first-class degrees rose from 8% in 1996-7 to 26% in 2016-7 and to a peak of 36% in 2020-1, before falling back to 30% in 2022-3.[6] As Figure 1.2 shows, this rise in the proportion of firsts was a general trend among English universities, applying especially to the research-intensive Russell Group, but also extending to the more professionally-focussed University Alliance and to other institutions. As the chart also shows, by the onset of COVID several institutions were awarding firsts to more than half of their graduating students.

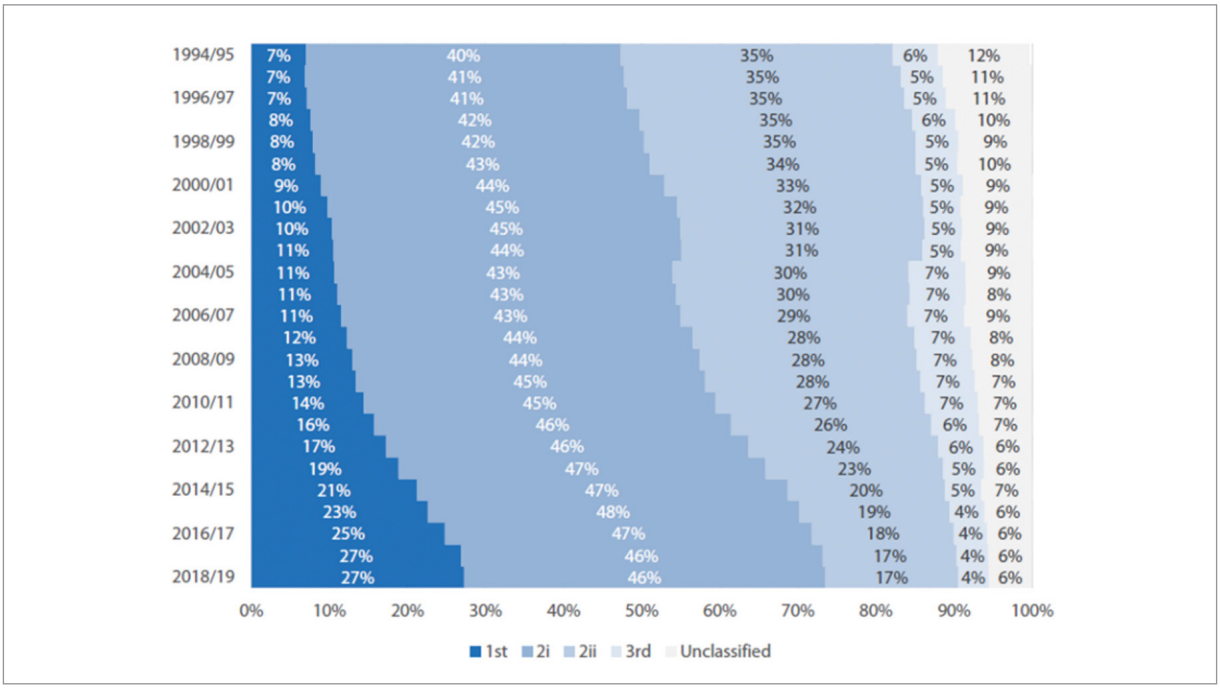
The proportion of 'good' degrees (a first or a 2:1) also rose substantially, from less than half of all degrees in 1994-5 to 80% in 2020-1, before declining slightly to 77% in 2022-3.[7] At Russell Group universities more than 91% of students received firsts or 2:1s in 2020-21, and at Oxford and Cambridge the figure was still above 90% in 2021-2.[8] By that same year, more than half of graduating students received a first or 2:1 at every university in the UK that awarded more than a hundred degrees.[9] Even by 2018, three quarters of all graduating students were receiving a 'good' degree, something less than half of their predecessors had achieved only 20 years earlier (as Figure 1.3 shows).[10]

In Canada, the proportion of A grades increased from 16% to 21% between 1973-4 and 1993-4 at a sample of Ontario universities. A's and Bs increased from 48% to 53%, and the total pass rate increased from 90% to 93%. Music awarded the highest proportion of As at both the beginning and end of the period (29% in 1973-4 and 43% in 1993-4).[11] The variance of the grades also decreased in most of the academic departments studied, and 'departments with the biggest decline in the variance of grades also had the biggest inflation.'[12] In other words, there was a tendency for grades to become more similar as they went up.

At the University of Calgary, the proportion of As and Bs increased from 60% to 69% in junior level (first- and second-year) courses and from 75% to 80% in senior (third-year and above) courses between 1992-3 and 2001-2.[13]

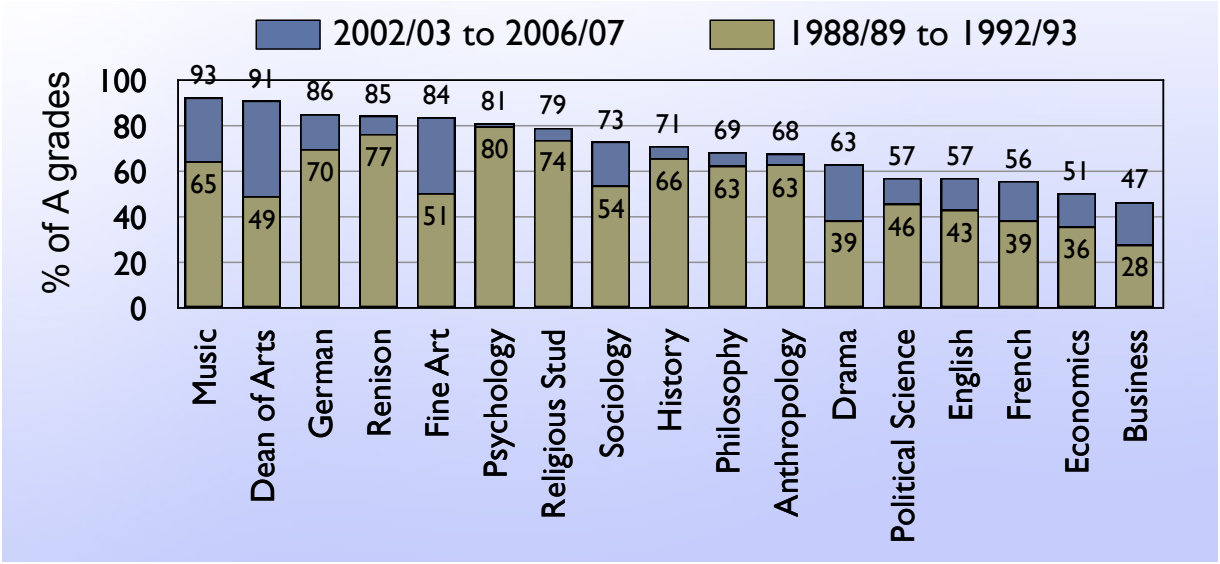
At the University of Waterloo, the proportion of A grades increased from 30% to 41% between 1988 and 2006. It also increased in every department at the university between the two periods of 1988-9 to 1992-3 and 2002-3 to 2006-7 (as Figure 1.4 shows). Music again awarded the highest proportion of A's (93%) in the later period.[14]

Figure 1.3: Classes of degree awarded at universities in England as a proportion of all degrees, 1994–2019



Source: 'UK grade inflation,' *Snippet Finance* (16 Sep 2020).

Figure 1.4: Average proportion of A-range grades awarded at the University of Waterloo in two periods (1988-9 to 1992-3 and 2002-3 to 2006-7)

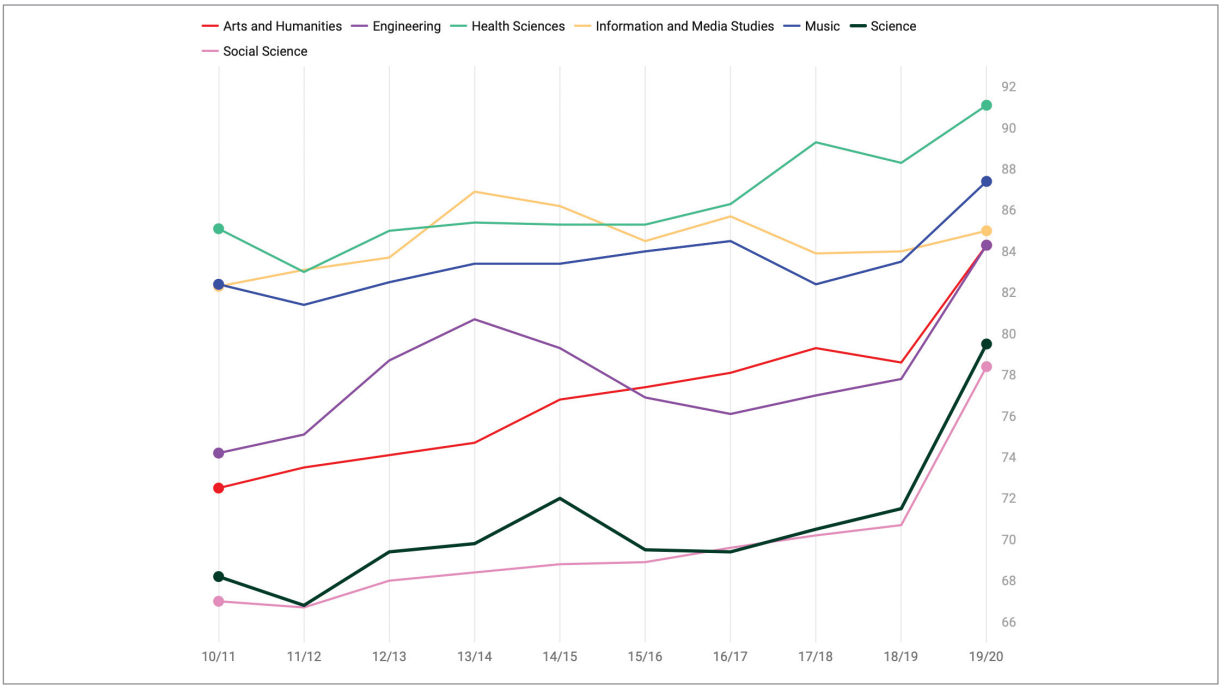


Source: Greg Mayer, "Grade Inflation," *op. cit.*

At the University of Western Ontario (as Figure 1.5 shows) the average mark increased in all faculties, including in Arts and Humanities (where it rose from 73% to 84%) and Science (where it increased from 68% to 80%) between 2010-1 and 2020-1. There was an especially steep rise in grades from 2019-20 to 2020-1, the first year of COVID measures.[15] At McGill between 2016 and 2021 the proportion of classes with averages in the A range increased from under 30% to 37%.[16]

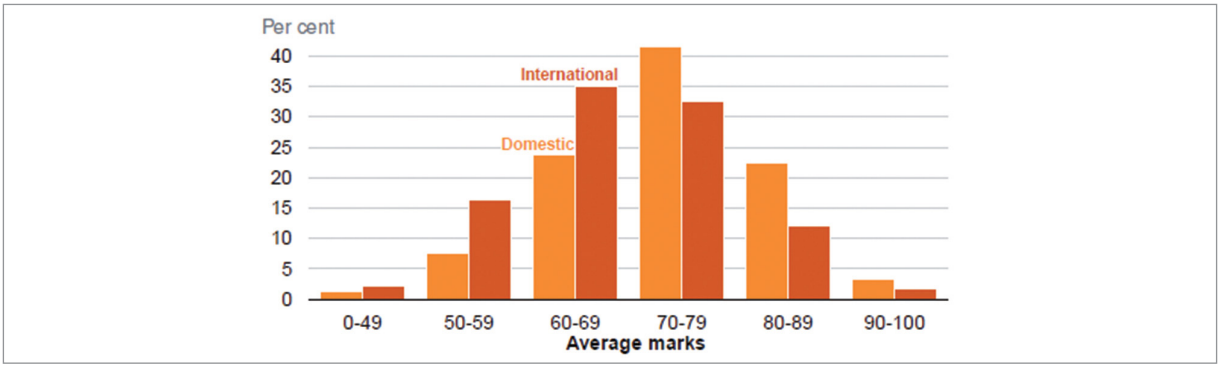
In Australia, an analysis of grades reported by students concluded that 'overall the distribution looks fairly normal, with most students in the middle,' even if 'maybe a quarter of domestic students reporting average marks above 80 per cent (see Figure 1.6 below) looks high.'[17]

Figure 1.5: Average grade by faculty at the University of Western Ontario, 2010-1 to 2019-20



Source: Cat Tang, "High marks," *op. cit.*

Figure 1.6: Average percentage marks for domestic and international Bachelor's students at Australian universities in 2016



Source: "Is grade inflation a problem?" *op. cit.*

Last year, though, a report by researchers at the University of Sydney showed that the proportion of high distinctions awarded at their institution increased from 7.7% to 25.7% between 2011 and 2021. The researchers controlled for students' performance during the final year of high school, the type of university degree they were pursuing, and their sex, finding that 'if two students who have the same ATAR [Australian Tertiary Admission Rank], go into the same degree and are the same gender,' then the 2021 student 'is going to get a mark 7.2 points higher' than his 2011 predecessor.[18]

And this year a survey of 100 academics at Australian universities found that 73% had seen grade inflation in their universities and that their main feelings about grade inflation were frustration (50% of respondents), powerlessness (44%) and dissatisfaction (31%).[19]

To sum up, there is good evidence grades have gone up in the US since the 1970s and in the UK since the 90s. In Canada there is evidence that grades have inflated at a few individual universities from the 70s until the COVID period, though how widespread the phenomenon has been is less clear. In Australia there is now solid evidence for grade inflation in the decade to 2021, at least at the University of Sydney.

What about New Zealand universities? Is grade inflation a problem here?

University grading systems

A number of different grading systems and degree classifications are in use across the English-speaking world.

UK universities issue three categories of degree: first class honours (1st), second class honours (2nd), and third class honours (3rd), with the second class being further divided into 'upper second-class honours' (2:1) and 'lower second-class honours' (2:2). These correspond to percentages as follows:

First-Class Honours (1st)	70% and above
Upper Second-Class Honours (2:1)	60–70%
Lower Second-Class Honours (2:2)	50–60%
Third-Class Honours (3rd)	40–50%
Fail	0–39%

When I was an undergraduate in the UK, we knew the various classes of degree by their monikers in (pseudo-)Cockney rhyming slang. We all wanted to get a Damien (Hirst or First) of course, though most of us would happily settle for an Elton (John or Two:one) – a class of degree that almost half of all graduating students attained in 2023. Some of us definitely worried about getting a Desmond (Tutu or Two:two), and we were terrified of going away with only a Douglas (Hurd or Third). Today's students are unlikely to harbour such fear of a Douglas, with only 3% of degrees awarded in 2023 being in the lowest passing category.

There is a good deal of variety across the almost 4,000 degree-granting institutions in the US, but most use letter grades, percentages, or a combination. Individual students are also given a GPA (Grade-Point Average) on the 4-point scale. Below is a fairly typical mapping of letter grades to percentages to GPAs.

Letter Grade	Percentage	GPA
A+	97–100%	4.0
A	93–96%	3.9
A-	90–92%	3.7
B+	87–89%	3.3
B	83–86%	3.0
B-	80–82%	2.7
C+	77–79%	2.3
C	73–76%	2.0
C-	70–72%	1.7
D+	67–69%	1.3
D	63–66%	1.0
D-	60–62%	0.7
F	0–59%	0.0

There is also a good deal of variation across Canada's over two hundred universities, but most Canadian universities also use some combination of letters grades, percentages, and GPAs, even if the mapping is often slightly different. A typical mapping is below.

Letter Grade	Percentage Range	Meaning
A+	90–100%	Exceptional
A	85–89%	Excellent
A-	80–84%	Very Good
B+	75–79%	Good
B	70–74%	Satisfactory
C+	65–69%	Average
C	60–64%	Below Average
D	50–59%	Pass
F	Below 50%	Fail

Australian universities award high distinctions, distinctions, credits, passes and fails, usually according to the following schema:

Grade	Percentage Range
High Distinction (HD)	85–100%
Distinction (D)	75–84%
Credit (CR)	65–74%
Pass (P)	50–64%
Fail (F)	Below 50%

New Zealand universities, finally, use a more conventional combination of letter grades and percentage marks. Below is how this schema appears on the Victoria University of Wellington website. This is the grading scheme I followed in my years as a lecturer.

Grade	Normal Mark Range	Mid-point	Indicative Characterisation
A+	90–100%	95	Outstanding performance
A	85–89%	87	Excellent performance
A-	80–84%	82	Excellent performance in most respects
B+	75–79%	77	Very good performance
B	70–74%	72	Good performance
B-	65–69%	67	Good performance overall, but some weaknesses
C+	60–64%	62	Satisfactory to good performance
C	55–59%	57	Satisfactory performance
C-	50–54%	52	Adequate evidence of learning
D	40–49%	45	Poor performance overall, some evidence of learning. Fail.
E	0–39%	20	Well below the required standard. Fail.

Grade inflation at New Zealand universities

As we saw at the beginning of this introduction, my own experiences as a lecturer at Victoria University of Wellington had led me to suspect that grade inflation is a problem at New Zealand universities.

And I'm not alone. Other academics and students (or former students) have had similar experiences. We will look at a few testimonies from academics and students in detail in Chapter 4, where they will help us think about what the drivers of grade inflations at our universities are.

In terms of establishing whether grades are going up, though, these anecdotes obviously aren't that helpful. Have there been any studies of grade inflation at New Zealand universities of the sort that we just reviewed for universities in the US, the UK, Canada, and Australia?

There has been sporadic interest in grade inflation. In 2001, the *New Zealand Herald* reported that a number of lecturers had said that grade inflation was a problem at their universities after one outgoing academic spoke out publicly on the topic.[20] And in 2018, an analysis of data from an Official Information Act request

revealed that a smaller proportion of grades were in the A range in undergraduate law courses at Victoria University of Wellington than at the universities of Auckland and Otago.[21]

There has not, however, been any systematic exploration of grade inflation using grades awarded at New Zealand universities over time. In this report, we propose to fill that gap.

Before we set out our roadmap for the rest of the report, though, we will need to broach three crucial issues. First, what is grade inflation? Second, how does it arise? And third, why is it a bad thing?

Grade inflation: what it is, how it arises, and why it's bad

The term 'grade inflation' is a metaphor drawn from the more familiar monetary inflation. In monetary inflation, prices increase or inflate, meaning that the same amount of money represents less actual value on the market. A five dollar note that may once have bought you a flat white no longer does. Readers can no doubt provide a host of other examples from the experience of the past few years.

Most economists agree that the main source of monetary inflation is an increase in the money supply over and above an increase in economic activity. For Milton Friedman, famously, inflation was 'always and everywhere a monetary phenomenon,'[22] that is, it is caused by governments printing too much money. If there are more dollars going around and the same amount of value in the economy, then more dollars will have to represent that value, and prices will rise.

The 'grade inflation' metaphor suggests that something similar may happen with academic grades. If universities hand out more and more As in a way that isn't justified by student performance, the value of an A will go down. The same job opportunities will 'cost' more As as As flood the market. Students who worked hard will see the value of their As decrease over time, just as workers in the economy see their savings decrease in value due to monetary inflation.

There are of course differences between monetary inflation and grade inflation. While central banks manage a country's money supply, grades are given out by thousands of academics on a class-by-class basis. While there is no upper limit to monetary inflation (something that has been tragically clear in periods of 'hyperinflation'), there is an upper limit to grade inflation, since once every student is getting an A+ (or 100%), no further inflation is possible. (This is the main reason for a phenomenon that was noted in one of the Canadian studies we surveyed above, of grades tending to become homogenous as they go up. The rising grades are 'compressed' against the natural ceiling of perfect student performance.) Assessment systems do sometimes respond to grade inflation by introducing higher grades, as when the A* grade was introduced in England's General Certificate of Secondary Education (GCSE) and Advanced Level (A-Level) exams in 1994 and 2010 respectively. But for most purposes grades have a hard cap (100% performance) that currencies lack.

Despite these differences, 'grade inflation' remains an apt metaphor. Understanding this fundamental metaphor also helps us arrive at a definition. Just as monetary inflation takes place when money is produced over and above economic growth, grade inflation occurs *when grades rise in a way that is not justified by improvements in student performance*. In other words, grade inflation is when grades go up but student learning decreases, remains the same, or rises less than grades are rising.[23] Strictly then, the fact that grades are rising is not sufficient, in itself, to prove that grade inflation is going on, since rising grades may simply be tracking rises in student achievement.

There are, after all, a good number of factors that may have led to improving student performance. Incoming students may be better than in the past. Tuition fees and state spending on universities have increased considerably in most English-speaking countries over the past few decades, as have the number of teaching assistants and ancillary staff. So has the proportion of female students, who have tended to perform better than males in recent years. Universities now give clearer guidelines about assessment than in the past. Some academics claim that norm-referenced assessment, in which students are compared to their peer-group, has largely given way to assessment against pre-determined standards, which give students credit for the level they have achieved regardless of their position in the class.[24]

Many studies have responded to this issue by attempting to control for factors of this sort.[25] The UK's Office for Students now prefers to speak of 'unexplained' rises in grades – that is, the amount of the observed rises that is not explained once other possible factors are taken into account.[26]

Grade inflation, to repeat our definition, is a rise in grades that cannot be justified by improvements in student performance.[27] And it arises when universities (or schools) give out too many high grades. We will look in detail at why institutions (and academics and teachers) are driven to distribute higher grades in Chapter 4. For now, though, we should turn to the final issue we wanted to address in this section: why grade inflation is harmful.

We've already mentioned one reason that grade inflation is harmful. Just as monetary inflation erodes the value of currencies, grade inflation erodes the value of grades. A's today are less scarce than they used to be, and so they are worth less. Students with straight As will be able to 'buy' less on the job market, i.e. they will have less access to the best jobs, because their As pack less of a punch than they used to in terms of impressing potential employers.

A related problem is that as top marks become more common, grades become more homogenous. (Remember that this is a feature of grade inflation but not of monetary inflation, since currencies lack the 'ceiling' that grades have.) When grades are more homogenous, their capacity to discriminate between different levels of performance is reduced. A middling student who has prioritized late-night gigs over early-morning lectures may come out with the same grade as a bright student who has worked extremely hard. That robs the industrious student of the fruits of their industry.

It also means that the less industrious student isn't exposed to the same incentives. If students know that their grades will be an accurate reflection of their work at university, then they will be more motivated to work hard and get the best out of their state-subsidized university education. If students have a sense, instead, that their true performance may be masked by generous grading, they will be less likely to hit the books with any fervour.

That, in turn, will mean that fewer students will be mastering the skills and knowledge that are required for a dynamic economy and a vibrant culture. New Zealand will perform less well relative to its peer nations on a range of fronts, and the overall quality of life will suffer.

There may also be distortions arising from differential grade inflation among different disciplines, something we will look at in this report. If students sense that they will get an easy A in their media studies courses, they are less likely to take more rigorous courses (in STEM, for example) that may give them more valuable skills – or even courses that they might find more intellectually fulfilling (in classics, say).

Finally, grade inflation may undermine the equality of opportunity that is a key component

of social justice. If the best grades are rare, students who do well at university will be more likely to see their hard work rewarded in the form of well-paying jobs and other life opportunities, whatever their socio-economic background. If the top grades mean less and fail to distinguish between excellent and mediocre students, on the other hand, social connections and other, less merit-based considerations may play more of a role in determining individuals' future salaries and life chances.

The roadmap for this report

In this report – following this introduction, which has set the scene and the terms of discussion – we will present evidence showing that grade inflation is a problem at New Zealand universities. We will ask what the drivers of grade inflation are in this country and suggest some possible remedies to the problem.

Our presentation of evidence will be divided into two chapters, Chapters 2 and 3. In Chapter 2, we will draw on grade distributions from all eight New Zealand universities to show that both the proportion of A-range grades and the pass rate have been rising for the past ten to twenty years, with a particularly dramatic increase in the COVID years. We will also examine which faculties have awarded the highest grades in recent times, and which faculties have seen the highest rates of grade inflation.

Since rising grades do not necessarily constitute grade inflation, in Chapter 3 we will consider a few factors that might have contributed to rising grades: the performance of incoming students, the proportion of female students, the level of university funding, and staff to student ratios. We will argue that none of these factors plausibly explain the magnitude of the increase in grades that we document in Chapter 2.

In Chapter 4 we will ask why grade inflation occurs, and why it has gotten so bad in this country. We will sketch out a simple model that view grade inflation as a collective action problem, and also draw on testimonies from New Zealand academics to gain more of an insight into what has been contributing to grade inflation here.

Finally, in Chapter 5 we consider how grade inflation might be reversed or at least mitigated in our universities. We consider various methods for moderating grades, reforms that would change the incentives behind grade inflation, and changes to the grading culture that is currently dominant.

In the end, it is unlikely that there will be a one-shot solution to grade inflation here or anywhere else. Some combination of policies may well make a positive difference, though, especially if it is founded on a wide-ranging and well-informed public debate on the issue. This is a debate that we hope to help stimulate with this report.

CHAPTER 2

Grades Are Rising

In this chapter we will lay out the evidence that grades have risen at New Zealand universities in recent years. This is the case both in terms of the proportion of grades in the A range (A+, A, and A-), and in terms of pass rates. In Chapter 3 we will ask whether this pattern of rising grades can be explained by factors other than grade inflation, such as improvements in student performance.

Data gathering and analysis

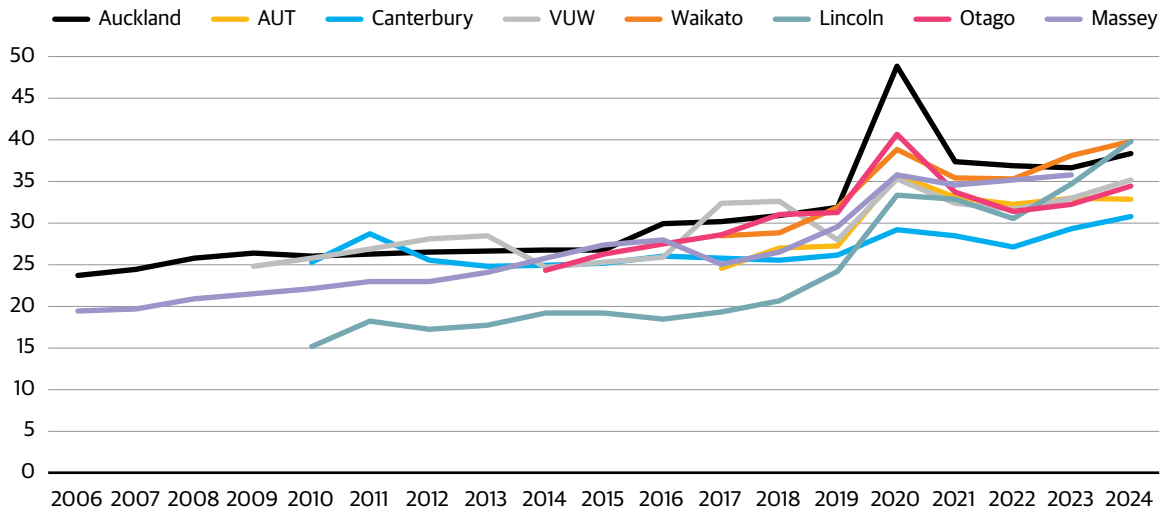
We asked all eight of New Zealand's universities to send us grade distributions for all undergraduate courses by year and by faculty stretching back as far as possible. In most cases we made our requests under the terms of the Official Information Act, which allows citizens to request information from public institutions. We received grade distributions from all the universities. Some went back as far as 2006 (Auckland, Massey). Others extended back only to 2017 (AUT, Waikato).[28]

A's and passes

As Figure 2.1 shows, the proportion of A-range grades as a percentage of total grades has gradually risen over the years.[29] The proportion of As rose steadily in the decade to 2019, then rose sharply during COVID. After COVID there was a brief adjustment downwards before the proportion started rising again.

Overall, the median proportion of A-grades grew by 13 percentage points, from 22% to 35%.[30] The largest increases occurred at Lincoln, where the proportion of As grew by 24 percentage points between 2010 and 2024 (from 15% to 39%), more than doubling, and Massey, where they grew by 17 percentage points (from 19% to 36%) from 2006 to 2023.

Figure 2.1: Percentage of A-range grades at New Zealand universities, 2006–2024[31]



Source: University records via OIA requests.

The fastest increases took place during COVID. From 2019 to 2020, A grades at Auckland went from making up almost a third (32%) of total grades to almost half (49%) of them. At Lincoln, an exceptionally low proportion of grades were in the A range until COVID, when the percentage of As shot up to levels comparable to other universities. By 2024, Lincoln had the joint-highest percentage of A grades in the country with Waikato (40%).

After COVID A grades fell but then rebounded, especially at Waikato, Lincoln and Canterbury, where they bounced back to unprecedented heights. But they also rose slightly at Massey, AUT, VUW and Otago, while at Auckland they rose slightly after a relatively low point in 2023.

Pass rates, depicted in Figure 2.2, show a more complicated pattern. At most New Zealand universities pass rates have clustered at around 90% for most of the past two decades. The only exception was Waikato, where pass rates were below 85% before and even during COVID, but which have since increased to over 90%.[32]

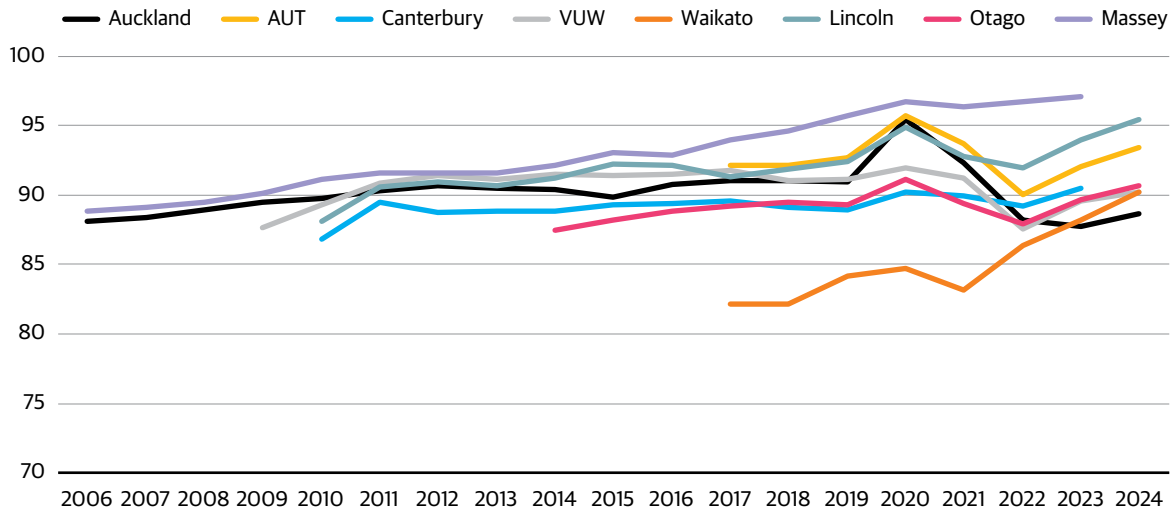
A spike in pass rates during COVID similar to the spike in As above is evident. At AUT and Auckland, this spike brought pass rates up to over 95%, while at Lincoln it brought them very near to that mark. Otago and Canterbury's pass rates also rose above 90% during the COVID period.

At Massey, pass rates tracked Auckland's into the 2010s, but then continued to rise steadily as Auckland's stabilised for a time. Massey's extraordinary trajectory meant that, although its pass rates hardly spiked during COVID, it still attained a 97% pass rate by 2020.[33]

At AUT, Otago, VUW and Auckland, pass rates fell in 2022 to lower levels than before COVID before once again heading upwards. (At Lincoln they fell to nearly the level of 2019 before a similar rebound.)

Unlike A grades, pass rates have long approached the natural ceiling of 100%, limiting their scope to increase. But the median pass rate at the median university did rise by 2 percentage points (from 88% to 90%) over the whole period, and no university's pass rate decreased in that time. The greatest increases were at Waikato and Massey, which both saw an 8 percentage point rise in their pass rates (from 82% to 90% and from 89% to 97% respectively).

Figure 2.2: Percentage of pass grades at New Zealand universities, 2006–2024



Source: University records via OIA requests.

Grading in different disciplines

We also examined which disciplines award the greatest proportions of A grades. To do this, we grouped the faculties for which we had grades into categories (which we will call 'disciplines,' even if some are somewhat artificial) as shown in Table 2.1.

Table 2.1: Our grouping of faculties at New Zealand universities by discipline

University	Arts & Humanities	Business & Economics	Creative Arts/ Design/ Architecture	Education	Engineering	Law	Health	Science
Auckland	Arts	Business and Economics	Creative arts and industries	Education and social work	Engineering	Law	Medical and Health Sciences	Science
AUT	Culture and Society	Business, Economics and Law	Design and Environmental Technologies				Health and Environmental Sciences	
Canterbury	Arts	Business		Education	Engineering	Law	Health	Science
Lincoln		Agriculture and Commerce	Faculty of Environment, Society and Design					Agriculture and Life Sciences
Otago	Humanities	Commerce					Health Sciences	Sciences
Massey	Humanities and Social Sciences	Business	Creative Arts				Health	Sciences
Waikato	Arts, Law, Psychology and Social Science	Management		Education			Health, Engineering, Computing and Science	
VUW	Humanities	Commerce	Architecture	Education	Engineering	Law		Science

Many of the choices we made in drawing up this table were self-evident. All the faculties of engineering clearly belonged in an 'Engineering' category, for example, and all the faculties of law belonged in a 'Law' category.[34] Arts and humanities faculties clearly belonged together, as did faculties of business and commerce. Placing Lincoln's Faculty of Agriculture and Life Sciences into our 'Science' category, and its Faculty of Agriculture and Commerce into our 'Business and Economics' category also seemed relatively unproblematic.

Then came some more difficult decisions. For composite faculties, such as Waikato's Faculty of Arts, Law, Psychology and Social Science, and its Faculty of Health, Engineering, Computing and Science, we simply allocated the whole faculty to the first category named – so to the category of 'Arts and Humanities' in the first case, and to the category of 'Health' in the second. We made a pragmatic decision to group faculties of creative arts, design, and architecture together in a single makeshift category of 'Creative Arts/Design/ Architecture.'

Special programmes

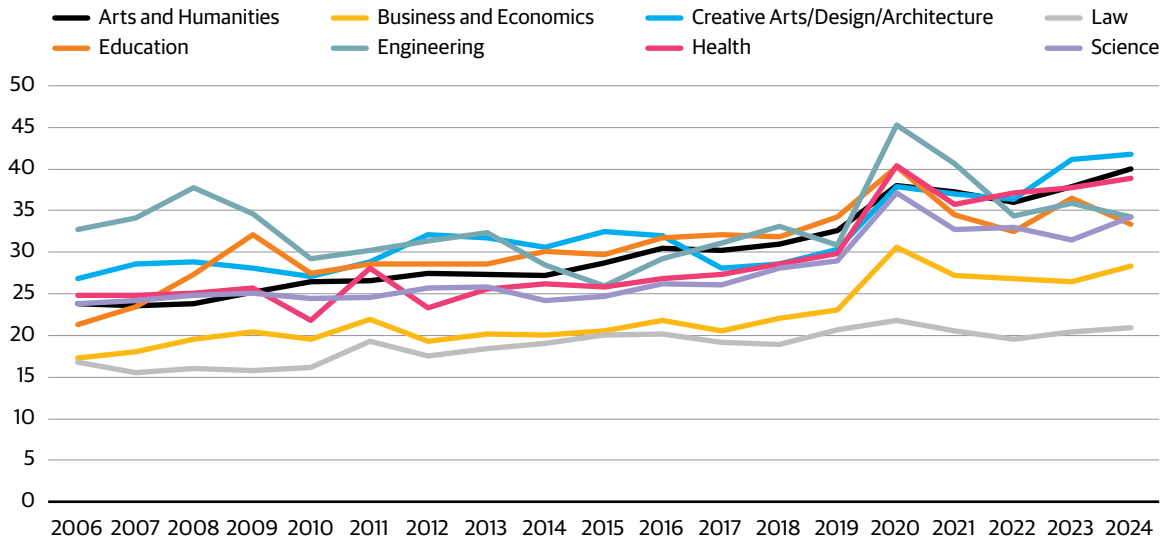
As well as more recognisable faculties, the grading distributions we were supplied also contained a small number of special faculties (University Programmes at AUT, University Studies and English Language at Lincoln, and Study Group at Waikato) and two small programmes that constituted their own faculties (Theology at Auckland and Māori Studies at Waikato). We excluded these from our graphs, as the small numbers of grades involved led to erratic results and might have given a misleading impression. It might be worth noting in passing, though, that these included the highest and lowest proportions of A grades that we came across while preparing this report.

The highest proportion of As that we met was the 66% of grades that were in the A range in University Programmes at AUT. When we asked AUT to tell us what 'University Programmes' was, Director of Communications Alison Sykora said it included programmes 'that don't "belong" to any one faculty,' that at AUT this largely refers to 'Certificates of Proficiency, which is when a student enrols in a single course rather than a full traditional programme such as a Bachelor of Arts,' and that 'it also includes our UniPrep programme.'

The lowest proportions of As we came across were in University Studies and English Language at Lincoln, where only 7% of the grades were As in 2010, although by 2020 the faculty had proportions of A grades that were more typical of other university faculties in New Zealand, at over 20%.

Figure 2.3 shows the median percentage of A grades awarded at New Zealand universities by discipline. Law faculties awarded the lowest percentage of A grades though the whole period, with the proportion of As consistently exceeding 20% only after COVID. Faculties of business and economics gave out the second lowest percentage of As; they, too, maintained this position for the whole period, with A grades exceeding 25% only after COVID. The other disciplines cluster together, rising from 20%–35% to 30%–45% after COVID.

Figure 2.3: Percentage of A-range grades at New Zealand universities by discipline, 2006-2024



Source: University records via OIA requests.

In all of the disciplines the percentage of As increased over the period. The smallest increases were in Engineering (just one percentage point, from 33% to 34%) and Law (4 percentage points, from 17% to 21%). The largest increases were in Arts and Humanities (16 percentage points, from 24% to 40%) and in Creative Arts/Design/Architecture (15 percentage points, from 27% to 42%).

We can also see the spike in grades during COVID that we noticed in Figures 2.1 and 2.2. As in those graphs, grades also rebounded after a fall following the COVID period. In Arts and Humanities and Creative Arts/Design/Architecture the proportion of As was even higher in 2024 than it was at its COVID peak. Every other discipline saw rises in the proportion of As after an immediate post-COVID fall.

But how much are our sometimes artificial categories influencing the results? To control for this, we compared just the faculties that did not include more than one discipline.[35] We also removed all the faculties in our Creative Arts/Design/Architecture category. The results, displayed in Figure 2.4, are very similar to those in Figure 2.3.

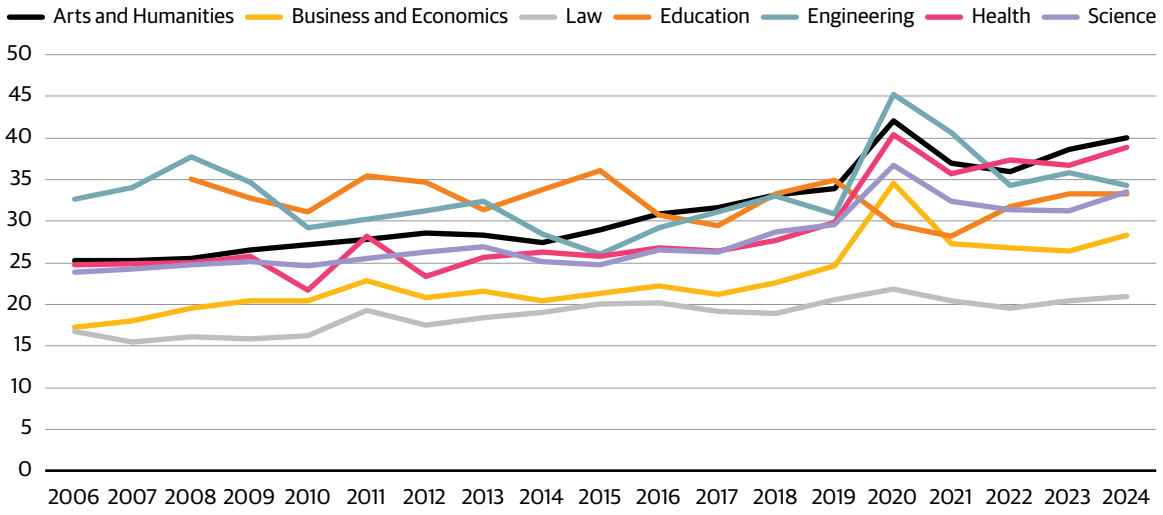
How do we compare with other countries?

In the introduction we surveyed grade increases in other university systems in the English-speaking world. How do the grade increases we have found at New Zealand universities compare to those?

The proportion of A grades at New Zealand universities was at a high of 35% in 2024, below the 45% figure that was observed at US universities in 2012. No New Zealand university awarded a percentage of As that was anywhere close to Harvard and Yale's figures (both just under 80%) in 2023, although an impressive 49% of grades awarded at Auckland in 2020 were in the A-range, and 40% of the grades awarded at Otago the same year were in the A-range.

The current proportion of A grades at New Zealand universities is about the same as the percentage of first-class degrees awarded in England at its peak in 2020-21 (36%). At the same time, the rise in the proportion of firsts was much sharper in England: the percentage of firsts there grew by 28 percentage points between 1996-7 and 2022-3, nearly trebling, whereas the percentage of A grades in New Zealand grew by 'only' 13 percentage points between 2006 and 2024.

Figure 2.4: Percentage of A-range grades at New Zealand universities by discipline, 2006–2024 (simplified)



Source: University records via OIA requests.

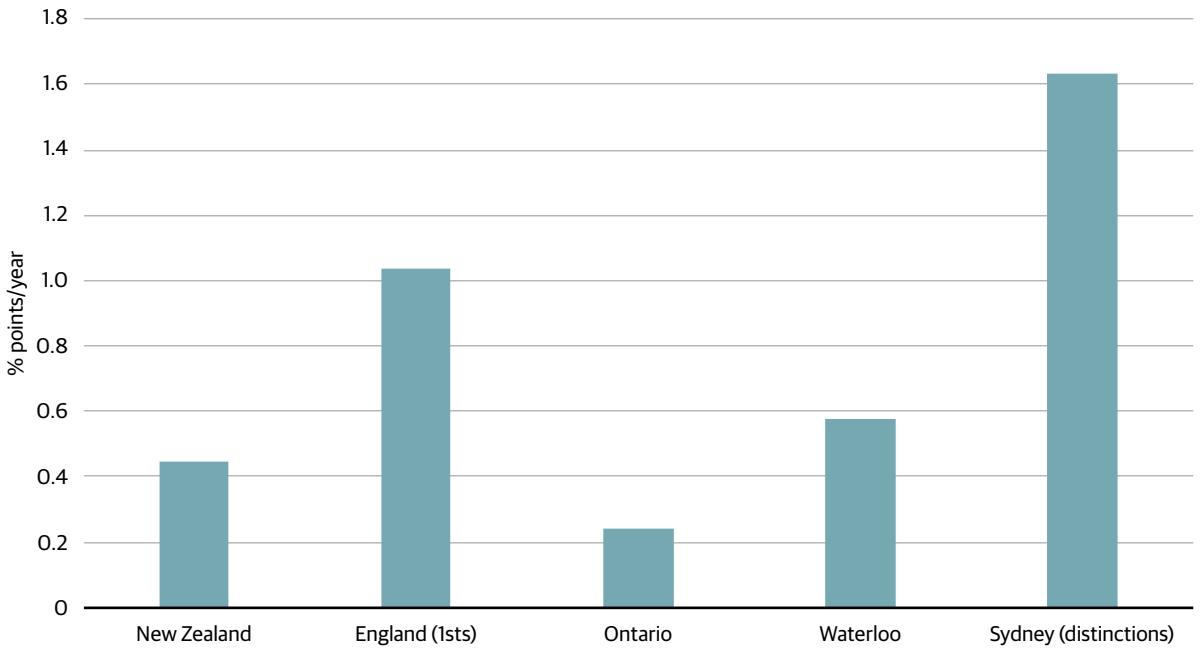
Of course, the measures here are different, with firsts being a classification awarded to entire degrees rather than a range of grades awarded in individual courses.

The median percentage of A grades at New Zealand universities was substantially higher than the proportion (21%) at a sample of Ontario universities in 1993-4, but lower than the percentage (41%) at one Ontario university, the University of Waterloo, in 2006. The 13 percentage point increase in the percentage of As in New Zealand, though, was much larger than the 5 percentage point rise at the sample of Ontario universities between 1973-4 and 1993-4, and slightly larger than the 11 percentage point rise at the University of Waterloo between 1988 and 2006.

Finally, the 13 percentage point growth in the percentage of As awarded at New Zealand universities was somewhat smaller than the 18 percentage point increase in the percentage of high distinctions awarded at the University of Sydney between 2011 and 2021. (We should bear in mind here that in Australia high distinctions are awarded for grades of 85% and above, whereas in New Zealand the A-range begins at 80%.)

We can calculate an annualized rate of change for these different periods by dividing the change in percentage points by the number of academic years. The results are displayed in Figure 2.5. below. We should again be mindful of differences between A grades on the one hand and first-class degrees and high distinctions on the other. But with these caveats in mind, the graph suggests that between 2006 and 2024, the proportion of As awarded at New Zealand universities rose at a rate that was considerably slower than the rise in the proportion of firsts in England and high distinctions at the University of Sydney in recent years, but was somewhat faster than the rise in the proportion of As at some Ontario universities (but not Waterloo) in the final decades of the 20th century.

Figure 2.5: Change in the percentage of As or other top grades awarded at universities in Canada, England, and Australia in percentage points per year, available periods



Sources: See the references in Chapter 1.

Summary and conclusion

To sum up, the percentage of A grades awarded at New Zealand universities rose by some 13 percentage points from 2006 to 2024, from 22% to 35%. The proportion of A grades spiked during COVID, with nearly half of all grades awarded at the University of Auckland in 2020 in the A range. The percentage of As fell after COVID measures eased, but then began rising again.

There was a similar pattern with pass rates. Here there was a much more modest increase, with the median rising only 2% from 88% to 90%. This may partly be a range compression effect; pass rates have always been much closer to the 100% ceiling than the percentage of A grades. There was a spike in pass rates during COVID, bringing several universities' pass rates above over 95%, then a decline immediately after COVID, and then a rebound, with pass rates heading upward again.

The percentage of As grew in all academic disciplines, with the largest increase being in in business and economics (12 percentage points) and the smallest in law (4 percentage points).

Comparing the rise in top grades at New Zealand universities to similar rises in other university systems is not always straightforward, as available studies sometimes use different grading conventions (like first-class degrees in England). They also cover different time spans. That said, if we calculate an annualized rate of change for the time series that we do have data for, the rise in the proportion of As at New Zealand universities looks slightly faster than past rises in the proportion of As at some Ontario universities (but not Waterloo), and considerably slower than recent rises in the proportion of firsts in England and in high distinctions at the University of Sydney.

CHAPTER 3

Why Are Grades Rising?

As we noted in Chapter 1, the fact that grades are rising doesn't necessarily mean that grade inflation is at work. Grade inflation, to repeat our definition, is when grades rise in a way that is not justified by improvements in student performance. Is it possible that the grade rises we observed in the previous chapter at New Zealand universities simply reflect better learning outcomes?

It is (just about) possible, but, as we will see in this chapter, it is extremely implausible. To establish this, we look at four factors that could theoretically have led to better learning outcomes in recent years. These are

1. an improvement in student performance in the final year of secondary school,
2. an increased proportion of female students (who tend to perform better than their male counterparts),
3. an increase in university revenues and expenditures, and
4. a decrease in the ratio of students to staff members.

All of these factors are now commonly taken into consideration in studies of grade inflation,[36] and we were able to obtain data for all of them. Other factors that are often taken into account, such as students' age and choice of courses, we lacked data for.

Our lack of rich data for individual students also limited our ability to produce models of the sort which often now feature in the most sophisticated grade inflation studies. We suggest that future research into grade inflation at New Zealand universities should seek to integrate individual student data into a model of this type.[37]

In what follows, we will instead appeal to basic logic. Where a variable – student performance at secondary school, for example – seems to have worsened rather than improved, can it possibly have led to an improvement in university grades? When a relevant variable has increased, has it increased at the same time as the grade rises we have observed were taking place (or recently enough that a lagged effect is realistic)? And when a variable has increased at the same time that grades were rising, was the increase substantial enough to explain the quite dramatic increases in grades we observed?

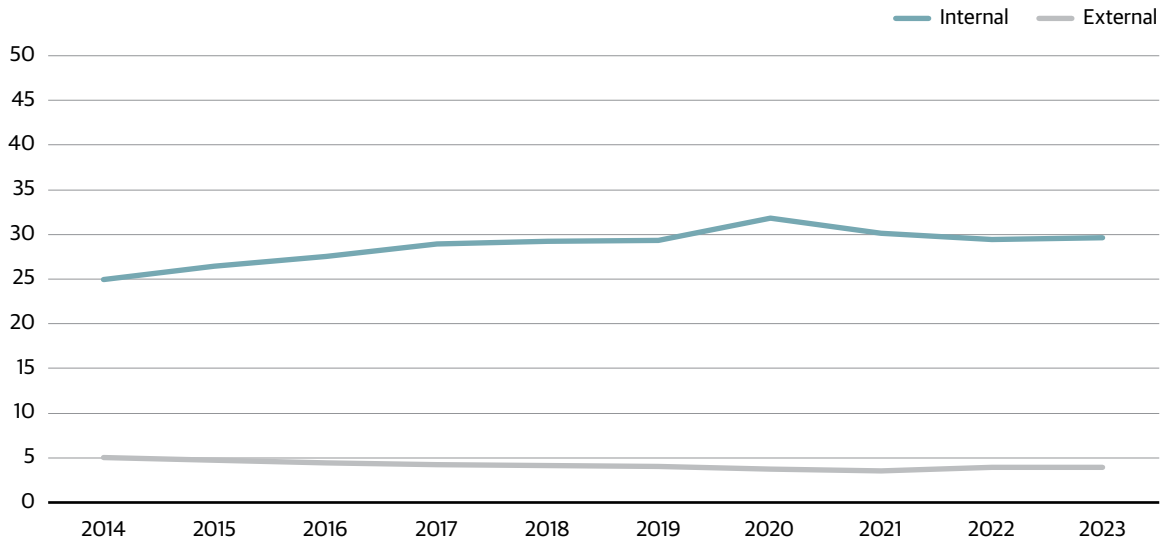
We will work through each of our four variables one after the other. In each case we will explain why the variable might plausibly lead to a rise in grades; show how that variable has changed over the past few years in New Zealand; and come to a conclusion about how much of a role (if any) changes in that variable could have played in the grade rises we described in the last chapter.

1. Better incoming students

The first possibility we want to examine is in many ways the most obvious: that university students are getting better grades because they were better students when they arrived at university.

To test this idea, it makes sense to look at student achievement during the final year of high school, when most students undertake NCEA Level 3. And it is results in the 'excellence' category that are the closest analogue for A-range grades at university level. Figure 3.1 shows the percentage of 'excellent' results at NCEA Level 3 from 2014 to 2023.

Figure 3.1: Percentage of results in the 'excellence' category at NCEA Level 3, 2014–2023



Source: New Zealand Qualifications Authority (NZQA) website.[38]

As the chart shows, the percentage of excellent results in internal assessments increased in this period, from some 25% to 30%. This seems to mirror the rise in the proportion of A-range grades that we observed in Chapter 2 (and displayed in Figure 2.1) from 21% to 35%. There is even a bump during COVID – not as dramatic as the COVID spike in Figure 2.1, but still observable.

However, there is reason to believe that NCEA internal assessments are themselves inflated. It is clear that they are not correlated with the externally assessed results displayed in the chart above. The percentage of excellence results in external assessments fell very slightly over the period, from about 5% to 4%. It was also always much lower than the percentage of excellent results in internal NCEA assessments, and than the percentage of A grades at universities.

The discrepancy is explained by the way in which each of internally assessed and externally assessed grades are allocated. Internal assessments are graded by students' own teachers, while externally assessed standards are graded by marking panels convened by the New Zealand Qualifications Authority (NZQA). The latter have technical mechanisms at their disposal to prevent grade inflation, which they routinely make use of.

On the other hand, while NZQA moderators check a proportion of internally assessed grades each year, the moderation is *post hoc* and incorrectly allocated grades are not adjusted. Rather, moderation is intended simply as feedback to improve teachers' marking the next time around.

Furthermore, when moderators disagree with teacher-allocated grades, it is in the downward direction in a vast majority of cases. The system for moderation of internally assessed standards simply exerts too weak an influence to control grade inflation. NZQA have been aware of the problem with grade inflation in internal assessment for some time. It was discussed by their technical advisory group as long ago as 2016.

Overall, NCEA performance (which mainly reflects internal assessments) has long been negatively correlated with New Zealand's scores in the international Programme for International Student Assessment (PISA). While attainment of NCEA Level 3 trended upward between

2008 and 2020, New Zealand's PISA results in reading, mathematics and science all trended downward during the same period.[39]

If we look just at the external NCEA results, it should be clear that the percentage of students who are reaching an outstanding level of academic achievement in their final year of high school has not been increasing. Indeed, irrespective of increases in excellence results for NCEA, the attainment rate for University Entrance (UE) has been essentially static since 2008.[40] This suggests that improved performance by incoming students cannot explain the rise in A grades that we observed in the last chapter.

2. More female students

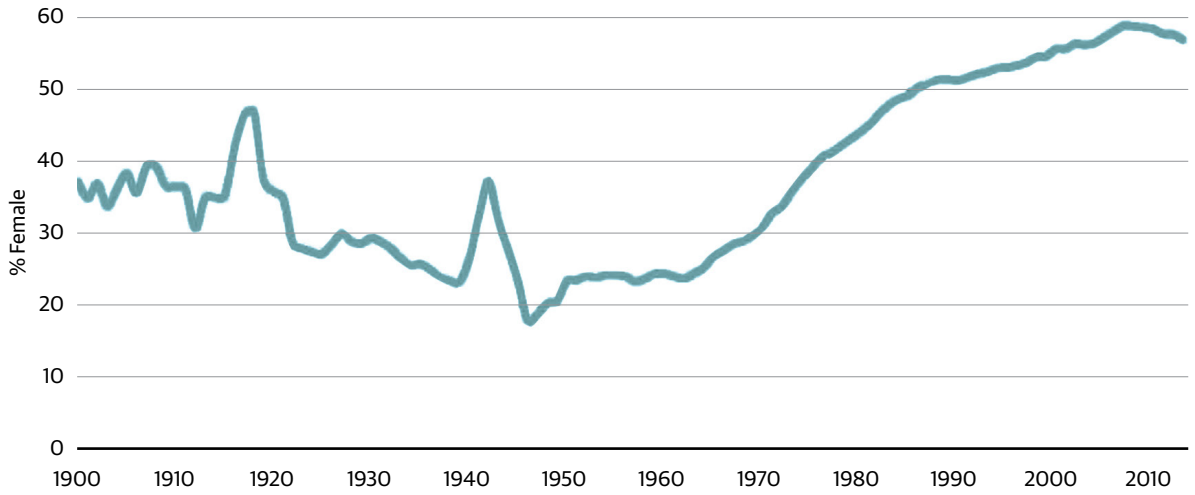
Women have had higher levels of attainment in tertiary education than men for some time now. Across OECD countries, 51% of 25–34 year-old women had a university degree in 2017, while only 38% of 25–34 year-old men did.[41] In the US, 46% of women had a degree in that age-bracket in 2021, while only 36% of men did.[42] In Australia, those figures were 52% for women and 38% for men in 2023.[43] Women made up 56% of Bachelor's degree holders in Canada in 2018, and 57% of higher education graduates in the UK last year.[44]

More women also get a 'good' degree in the UK than men, with 30% and 49% of women obtaining a first or a 2:1 respectively in 2024 compared with 27% and 47% of men.[45]

Women outperforming men at undergraduate level is now so routine that greater female participation is regularly considered in studies of grade inflation among the possible causes of rising grades.[46] (Note that it doesn't matter here why women tend to receive higher grades than men at university. The fact that they do – for whatever reason – means that greater female participation should lead to higher grades all round.)[47]

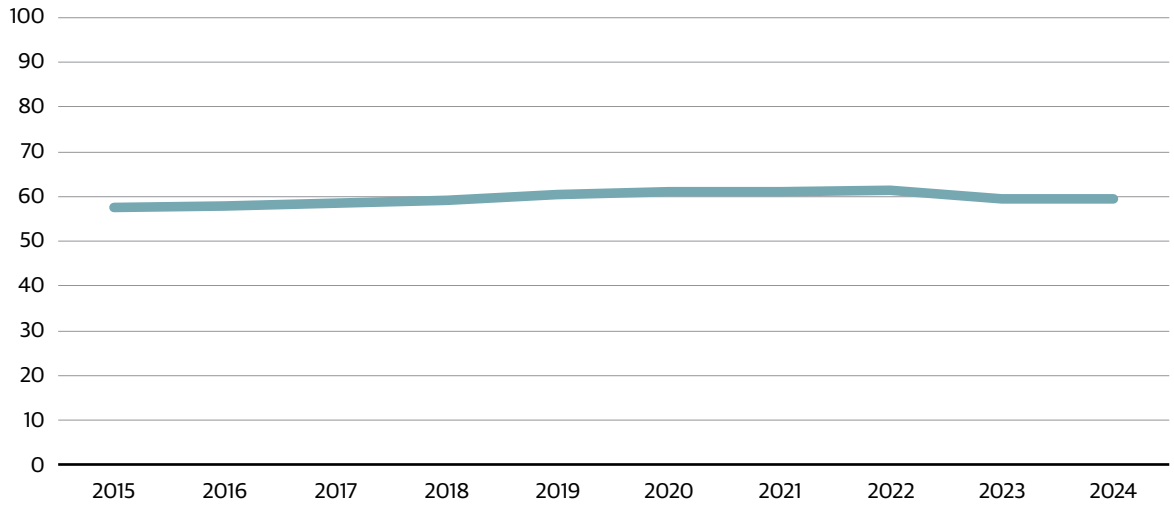
Figure 3.2 shows the percentage of women enrolled at New Zealand universities from 1900 to 2015.[48] As the graph shows, after a decline in the middle of the twentieth century, the proportion of female students began to grow steadily in the 1960s, passing the 50% mark in the 1980s and almost hitting 60% in the late 00s.

Figure 3.2: Women as a percentage of total enrolments at New Zealand universities, 1900–2015



Source: Adapted from Paul Stock *et al.*, "Female Student Participation," *op. cit.*

Figure 3.3: Percentage of female enrolments at New Zealand universities, 2015–2024



Source: Ministry of Education *Education Counts* website.[49]

Figure 3.3 shows the percentage of female enrolments at New Zealand universities since 2015. It shows that the percentage of female students passed the 60% mark in 2019 but then dipped back below it in 2023 and 2024. Importantly for our purposes, though, the proportion of female students has varied only within a four-percentage point band since 2015. Female participation has been essentially static over the last decade.

This presents a major problem for anyone looking to explain the rise in the proportion of As at New Zealand universities by appealing to a rise in the proportion of female students.

We should recall that the proportion of A grades at the median New Zealand university increased by 13 percentage points between 2006 and 2024. While there was an increase in the proportion of female students during that period, this was from just under to just over 60%, and seems unlikely to have played a major role in the substantial increase in A grades over the same period.

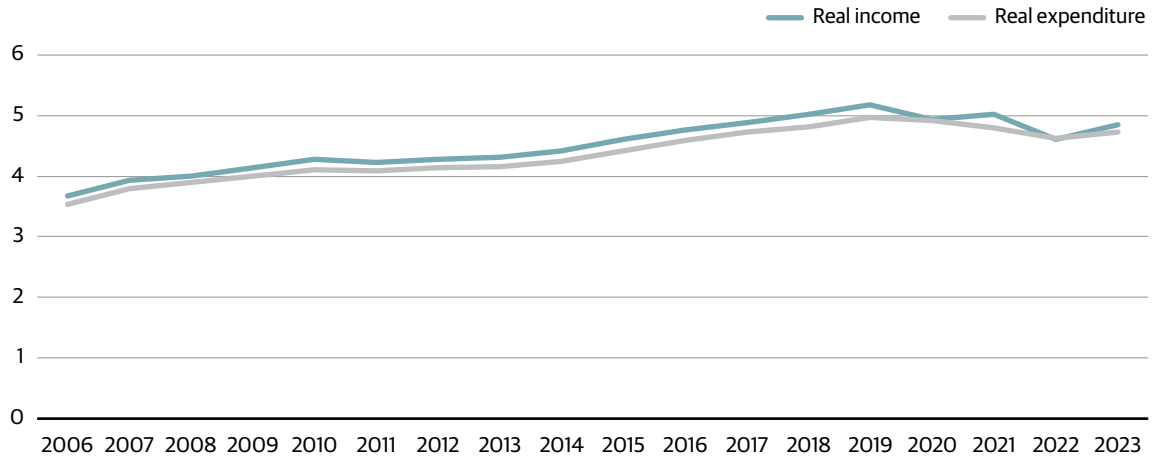
The much more substantial increase in the proportion of female students between the 60s and the 00s took place before our data on grade distributions begins. It cannot, therefore, explain the rises in grades we observed in Chapter 2.

3. Mo' money

Another possibility is that better funding has led to improvements in student performance. This might seem intuitive to some, although whether increased funding in itself leads to better educational outcomes is controversial.[50] A few US studies have found that cutting funding at universities leads to lower rates of degree completion.[51] It is certainly a possibility worth considering, then, that increased funding might have led to better student achievement at universities in this country.

Figure 3.4 shows the income and expenditure of New Zealand's universities in real terms (that is, adjusted for inflation) from 2006–2023.

Figure 3.4: Inflation-adjusted income and expenditure of New Zealand universities in billions of New Zealand dollars, 2006–2023



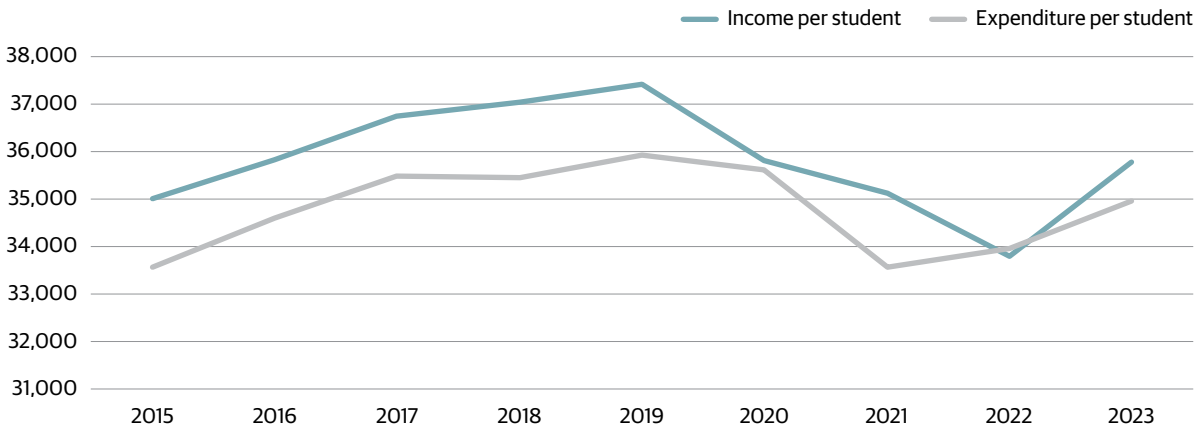
Source: Tertiary Education Commission (nominal income and expenditure figures) and World Bank (consumer price index figures) websites.[52]

As the graph shows, New Zealand universities' income and expenditure slowly increased from 2006, when both were less than $4 billion, to 2019, when universities' total income passed the $5 billion mark and their total expenditure peaked at just below it.

It may be, though, that student numbers changed substantially during this time, so that the graph above obscures a growth in funding per student. And if our intuition is that more money will lead to better educational outcomes through more staff, better resources and so on, it makes sense to see how much money has been put into the system per student.

Figure 3.5 shows universities' income and expenditure divided by the number of students at universities. As we can see, both income and expenditure per student rose until 2019, at which point they both dramatically declined before rebounding to some extent.

Figure 3.5: Inflation-adjusted income and expenditure of New Zealand universities per student in New Zealand dollars, 2006–2023



Source: Sources for previous figure plus Ministry of Education.[53]

To what extent do these graphs support the idea that the grade rises we observed in Chapter 2 can be explained by increases in university funding?

Here it may be useful to look back at our graph of the proportion of A grades at New Zealand universities over the same period, which we re-print as Figure 3.6 below.

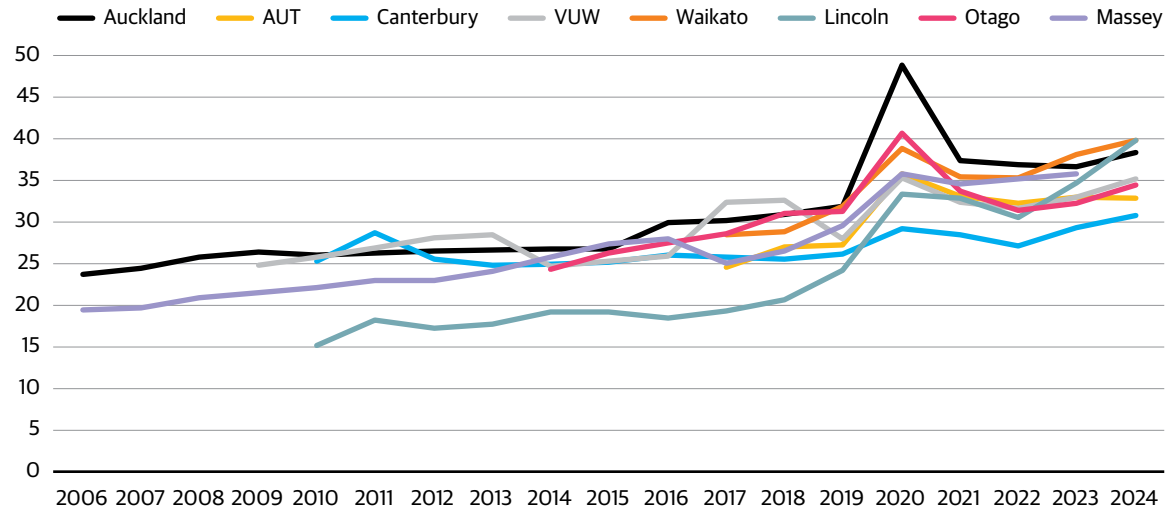
The rise in funding and expenditure from 2006 to 2019 in both absolute and per student terms does seem to be mirrored in rising proportions of A grades over the same period. And the increase in university spending from $35 to $50 billion (or from $35,000 or less to considerably more than $35,000 per student) might also seem substantial enough to explain all, or part of, the median increase in A grades from 22% to 35% of total grades awarded.

If the universities' financial positions are what explains all or most of the rise in grades, though, it is very difficult to explain what happens after 2019. From 2019 to 2023 university income and expenditure fell to $48 billion and $47 respectively in absolute terms – a fact often pointed to by academics appealing for more funds.[54] Income and expenditure per student also fell, even accounting for the rebound at the end of the period.

As Figure 3.6 reminds us, though, the proportion of A grades skyrocketed during this period. Just as university revenues and spending were declining slightly in absolute terms and collapsing in per student terms, grades were surging upwards at unprecedented rates. This suggests that increases in funding do not explain the overall pattern of grade rises we observed in Chapter 2.

Most people probably think it's plausible that more money would lead to better learning partly because money can buy better resources (more books in the library, better computers and IT support, and so on). The main reason the idea sounds intuitive, though, is because more money can buy more staff. So let's now turn to our final plausible reason that student performance might have improved: more staff.

Figure 3.6: Percentage of A-range grades at New Zealand universities, 2006–2024



Source: University records via OIA requests.

4. More staff per student

There are several possible reasons why we might think that more staff per student would improve educational outcomes. More teaching staff per student usually means smaller classes, in which instructors can give more personalized attention and feedback to individual learners. More non-academic staff per student might mean better access to various types of support, including learning support. Both these things might well lead to better student performance and hence to justifiably higher grades.

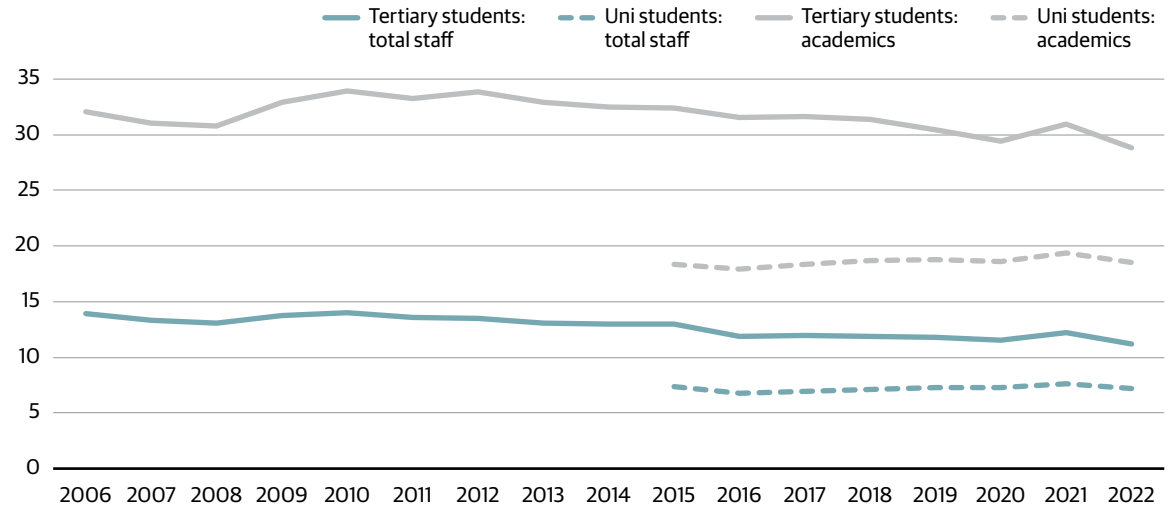
As with the idea that more funding leads to improved learning, the idea that smaller classes lead to better educational outcomes seems intuitive, but has long been controversial within the educational research community.[55] Some recent studies have found that smaller classes do lead to better outcomes, if perhaps only once classes are already smaller than about thirty students.[56] As for support staff, a recent study found that UK universities that moderately increased their proportion of non-academic staff had higher levels of degree completion but not of research quality, the number of 'good' degrees, or graduate employability.[57]

All in all, then, the idea that more teaching and/or support staff per student leads to better learning outcomes is one worth considering. Could this have been behind the rises in grades we observed in the last chapter?

Figure 3.7 shows staff-to-student ratios from 2006 to 2022. Since only total tertiary student numbers were available on the Ministry of Education website up to 2015, the graph only displays ratios of tertiary students to university academics and support staff up to that date. After 2015, it shows ratios of university students specifically to university academic and 'other' (non-academic) staff.[58] The similarities between the lines after 2015 might suggest that the ratios before 2015 are not too far wrong.

As the graph shows, the number of tertiary students per academic staff member declined from 32 to 29 students per academic. The number of tertiary students per staff member (academics and other staff) also declined, from 14 to 11 students per staff member. These declines could conceivably have meant more attention and support for each student, and thus could play a role in explaining the grade rises we saw in Chapter 2.

Figure 3.7: Students per staff member (FTEs) at New Zealand universities, 2006–2022



Source: Ministry of Education.[59]

At the same time, the declines in the student to staff ratio are quite small, amounting to only about three fewer students per staff member over the entire 16-year period. Again, we have to ask how much of the 15-percentage-point change in the proportion of A grades we think this could explain.

More seriously, when we get to the ratios of university students to staff – the most important ratios for our purposes – we do not observe any decline. Instead, the number of students per staff member is the same in 2021 as it was in 2015: 18 per academic staff and seven for academic plus non-academic.

We should recall that this is in a period when (as Figure 3.6 reminds us) the proportion of A grades increased rapidly, especially during the COVID years. That we don't observe any rapid decrease in the student to staff ratio in these years suggests that the number of staff per students is not having any important impact on the proportion of A grades.

This seems to be confirmed when we consider the small spikes in the student to staff ratios in 2021. These make clear that the number of students per staff actually increases around the time that the proportion of A grades skyrockets. That runs directly counter to the hypothesis that lower student to staff ratios (or more staff members per student) might help explain the grade rises we saw in the last chapter.

Conclusion

To repeat ourselves once again, grade rises do not necessarily constitute grade inflation. Grade inflation is when grades rise without a corresponding improvement in student performance.

To see whether student performance has improved, we might be tempted to turn to grades. But, of course, if we are examining whether grade inflation has occurred, we can't assume grades are a reliable indicator of student performance!

In view of this, the best recent research has looked for other factors that could have led to improvements in student performance. Since it's hard to measure student performance directly (especially if we can't assume that grading is reliable), we have to turn to variables that could plausibly be associated with better performance.

In this chapter, we have looked at four factors that are often looked at in studies on grade inflation.

We found that an improved student intake probably wasn't the reason that university grades have risen, since grades in the external NCEA assessments in the final year of school have not improved over the last few years. (Grades in internal assessments have risen, but these are almost certainly themselves inflated.)

We found that more female students (who tend to perform better) also probably wasn't the reason that grades have risen. Female participation rates did indeed surge from the 1960s to the first decade of the 21st century, but since then they have stagnated, meaning that they cannot explain the rise in the proportion of As over the past couple of decades.

More funding probably wasn't the reason that grades have risen either. Universities' incomes and expenditures did increase to 2019 (both in absolute and per student terms), but fell back during the COVID years, just as the proportion of A grades was skyrocketing. That would be hard to explain if more money led directly to better grades.

Finally, we saw that more staff members (academics or support staff) probably wasn't the reason that grades have risen. The number of staff per students at universities didn't rise at all during COVID – the period when A grades went

through the roof – which means that more staff members per student probably doesn't lead to better grades in any straightforward way.

Of course, this is not an exhaustive list of factors that could have led to improvements in student performance. But these are four of the factors that are most commonly appealed to by those arguing that rises in grades don't entail grade inflation in particular cases. In this particular case, we have seen that these factors probably do not explain the rises in grades.

Other researchers may want to propose other factors that may explain the rise in grades at New Zealand universities without us having to appeal to grade inflation. We would welcome proposals and analyses of this nature. In the meantime, we can say that we are not currently aware of any other factor that has been shown to have impacted the change in grade distributions in this country.

That means that we can now adopt as our working hypothesis in the rest of this report that the grade rises we have charted probably do constitute grade inflation. And that we can now turn to the question of how grade inflation arises in general, and how it may have arisen in our universities over the last couple decades.

CHAPTER 4
How Did We Get Here?

In Chapter 2, we saw that grades have gone up in New Zealand universities over the past two decades. In Chapter 3, we saw that these rises probably can't be explained by improvements in student performance. Since grade inflation by our definition is when grades rise without a corresponding improvement in student performance, it follows that grade inflation has probably taken place at New Zealand universities over the last twenty or so years.

In this chapter we will try to answer the natural next question: 'Why has this grade inflation taken place?' In Chapter 1 we noted that just as monetary inflation results from governments printing too much money, grade inflation results from instructors giving out too many high grades. But this obviously isn't anywhere close to a complete explanation for the phenomenon or for why it has become a feature of New Zealand universities.

In this chapter, we will present a simple model of how grade inflation occurs. In this model, grade inflation is a collective action problem in which behaviour that makes sense at the level of the individual academic (or even of the university) produces results that are sub-optimal for the university sector and for society as a whole.

We will also draw on testimonies from academics and students to get a sense of the incentives that are driving grade inflation at New Zealand universities. Finally, we will try to explain the striking spike in grades during COVID.

The grade inflation game

In 1968, the ecologist Garrett Hardin published an essay entitled 'The Tragedy of the Commons.' In it, Hardin drew on a scenario that had also been employed by the Victorian economist William Forster Lloyd. The scenario was as follows.

Imagine a village with a central 'commons' on which anyone can graze their animals. For an individual villager, it makes sense to let as many of his animals graze on the land as possible. But if every villager does that, the pasture will be depleted, and nobody's animals will be able to graze.

The extent to which the scenario Hardin presented reflects any historical reality has been debated. So have Hardin's precise description of the scenario and his recommendations for how to mitigate it.[60]

Nonetheless, Hardin's 'tragedy of the commons' clearly captures a common type of problem in human societies. What seems good for each individual can sometimes end up having negative consequences for everyone – including, of course, each individual concerned.

The tragedy of the commons is only one of a number of so-called 'collective action problems' that have been studied by economists and other social scientists.[61] Collective action problems often involve courses of action that are rational for individuals in the short term but can have adverse consequences over the long term if lots of people pursue them.

It is our contention that grade inflation is the product of a dynamic that is not dissimilar to the tragedy of the commons. Just like Hardin's villagers, academics pursue a good (in this case high student numbers) in a rational way (in this

case by awarding more high grades). And just as with Hardin's villagers, negative consequences ensue, with a common resource (sound grading) being depleted, to the cost of every individual academic as well as others.

In what follows, we sketch out a simple model of the incentives that we think lead to grade inflation at New Zealand universities. With a nod to game theory (a branch of mathematics which has shed a good deal of light on collective action problems), we call our model 'the grade inflation game.'[62]

In the grade inflation game, the good that academics want to maximize is student numbers. Individual academics, on the whole, want to have as many students in their courses as possible. This suggests that they are popular teachers and can help get them promoted (and hence gain more money and prestige). It can also help make sure the courses they want to teach stay on the menu.

The players of the grade inflation game are academics (including those in administrative roles). Academics generally assume that giving out easier grades will maintain or increase their student numbers. They are aware that other academics are also going to be gradually giving out higher and higher grades, which means that they will also have to follow suit just to keep the same numbers of students. And they are aware that they will have to grade more liberally than average to stand any chance of substantially increasing their enrollments.

Most academics also want positive student feedback, especially on the official feedback forms. Student feedback plays a key role in how academics are assessed for promotion and in applications for jobs at different institutions. It can also lead to particular courses and instructors having a good reputation within the student body, which can in turn support student numbers. And as a number of studies have shown, giving out higher grades leads to better student feedback.[63]

Academics who are inflating grades could theoretically face negative incentives at the programme level, either from their head of programme or from colleagues at end-of-term moderation meetings. But programmes have very little reason to come down strongly on grade inflation. Their funding, after all, is allocated according to student numbers, and that means that plummeting or even gradually declining student numbers could lead to layoffs – that is, to people losing their jobs.

The next possible checkpoint is the school – the School of Languages and Cultures in which I used to work at Victoria University, for example. Here again, heads of school could compel academics to grade to a curve, and school-level committees on teaching could enforce revisions to grades. Like programmes, though, schools also have an interest in attracting as many students as possible. And that means that they aren't likely to push back very strongly on grade inflation either.

The same incentives apply to faculties and even entire universities. Faculties compete for students with other faculties, both within and across universities, and universities compete for students with other universities. And all faculties and all universities would rather have more students rather than fewer, because more students mean more revenue.

Of course, a lot may go into which school, programme, or faculty students take courses in besides how easily those entities grade, and the choice of which university to attend is similarly complex. But studies have shown that grading practices do have a significant influence on which courses students choose.[64]

And all our model assumes is that at every level of academia, enough academics think that harsh grading risks losing them students, and that

generous grading might gain them students, that this has an effect on what grades are awarded in aggregate.

Academics then, can be compared to the villagers in Hardin's tragedy of the commons, since they, like the villagers, pursue their private good in the short term by handing out better and better grades in the hope of maintaining or increasing student numbers. But what is the commons in the grade inflation game? What is the common resource that is depleted when academic inflate grades?

Drawing again on monetary theory, we call this common good 'sound grading' (on the model of 'sound money.')[65] (For more on why we consider this to be a *common* good, see the text box.) Like sound money, sound grading means that currency units (here, marks or grades) maintain their power to signal value. Sound grading preserves its signaling power because it is reliable (grades correlate with actual performance), well-calibrated (the distribution of grades reflects the distribution of performance) and valid (grades measure what they are supposed to measure), and because it maintains these qualities over time.

What kind of good does grade inflation undermine?

Social scientists who study collective action make two main distinctions with regards to the kind of goods involved.

The first is between *excludable* or *non-excludable* goods. An excludable good is one that I can easily prevent others from consuming, like a mandarin that I can prevent my colleagues from eating by putting in my desk drawer. A non-excludable good is one that I can't easily prevent others from consuming. One example is clean air, that I can't easily prevent people from breathing.

The second distinction is between *rivalrous* and *non-rivalrous* goods. If a good is rivalrous, my consuming some of it will mean you will consume less of it. A mandarin can also serve as an example here: if I eat half of it, you will have to make do with half. By contrast, no matter how much air I breathe in, the amount you breathe in will not be significantly reduced. That makes air non-rivalrous as well as non-excludable.

Unlike mandarins or air, some goods are excludable and non-rivalrous, or rivalrous or non-excludable. The table below shows the various possibilities, and names economists have given to these types of goods.

	Excludable	**Non-excludable**
Rivalrous	private goods	common goods
Non-rivalrous	club goods	public goods

We have already come across an example of a good which is both excludable and rivalrous: a mandarin. A mandarin is thus a private good. We have also come across an example of a good which is non-excludable and non-rivalrous: air, which is therefore a public good.

A club good (a concept developed by economist James M. Buchanan) is one that I can exclude others from but which anyone not excluded from can use without reducing the amount available to others. A contemporary example might be movies and TV shows on Netflix.

A good example of a common good is a commons. Villagers can't easily stop one another from grazing on it, but one villager grazing his animals will deplete what other villagers' animals can graze on.

We would suggest that one of the ways in which grade inflation is like the tragedy of the commons is that the good involved (sound grading) is a common good. Academics cannot easily prevent one another from drawing on its authority while issuing high grades. But academics continuing to issue A grades do slightly deplete others' ability to issue credible As.

Sound grading is a valuable resource for society for several reasons. It allows students to have an accurate sense of where they are academically, helping motivate them. Smart students who work hard can be assured that their grit will pay off in good grades that not everyone will have achieved. And less talented students will be motivated to work as hard as they can by the knowledge that they will be awarded a poor grade if their work doesn't reach an acceptable standard.

The reliable sense that students have of their academic performance will also allow them to make more evidence-based decisions about what they should study. Sound grading also allows employers to identify the best students more quickly and reliably, allowing them to minimise time-consuming assessment processes.

All of these processes produce further benefits for society as a whole. Students who have been better motivated by sound grading will become more capable graduates, and more capable graduates will mean more productive workers across a whole range of different economic sectors, from technology to the arts. It should also mean higher-quality public debate, and a country that is better equipped to make good decisions on the issues that face it.

So students, employers and the public do benefit from sound grading. In our model, though, it is mainly academics that draw on the common good of sound grading. Academics take advantage of the reputation that high grades still have for signaling excellence, and they exploit this by giving them out cheaply in return for students attending their classes. But as with governments that print money in order to buy what they want – whether that's guns, butter or (indirectly) votes – the act of producing and giving out more high grades simultaneously slightly undercuts their value.

So it is that academics, to return to Hardin's commons, deplete a common resource by pursuing their private, short-term good. University students, in this scenario, are like cows or sheep that academics can have more of by providing them with more of a common resource (grass or grades).

Note, finally, that as with Hardin's villagers, though academics in our grade inflation game are initially imposing costs on students, employers, and on society as a whole, over the long term there should be costs for them as well. This is not only because academics are part of society and so stand to lose from a less educated population, a less dynamic economy, and a less vibrant artistic scene. It is also because ultimately it seems inevitable that the penny will drop and employers and the public at large will realize that university grades (and even university degrees) are no longer a reliable indication that someone is intelligent, industrious, or well-educated. Unless, of course, that process has begun already.

The grade inflation game at our universities

The grade inflation game that we just described is our best approximation to the dynamics that we believe are fueling grade inflation at New Zealand universities. In the rest of this chapter, we will draw on our own experiences as well as on anonymous testimonies from New Zealand academics to get more of a sense on the incentives that are operating on instructors.

First of all, we should be clear that funding based on the number of equivalent full-time students (EFTSs) has very much been a reality since the market-oriented reforms of the early 90s.[66] I have a strong sense of this personally, since Victoria University justified its 2023 cuts on the basis of falls in student numbers. Programmes with low student numbers were made to shed staff, since fewer students meant less funding.

From my own experience at Victoria I can also say that the belief that harsh grading could

turn students off courses and programmes was widespread. This was the point of the telling-off I received at the end of my first term at Vic. The implication was clearly that unless I graded less harshly, I would be losing students not only for my courses, but also for the programme.

A University of Auckland science lecturer described this virtually universal fear that tough grading will lead to fewer students and hence to less funding:

> Everyone is nervous about keeping students in their university and program, because administrations have made it pretty clear that students = funding = FTEs [full-time equivalents. i.e. academic positions]. If your programme is perceived as too tough, your programme may lose students to other programmes, i.e. you are sawing off the branch you are sitting on.

One thing I remember being told in my first year as a lecturer was that I could not, under any circumstances, fail more than 20% of the students in any one course, since (I was told) this would lead to funding for the course being terminated.[67] While this didn't present any problems in my Greek and Latin classes, which tended to have small numbers of excellent students, it was a concern in my large first-year lecture course, where performance in the bottom third of the distribution was very poor indeed.

In fact, so many students in this course would simply disappear and not hand in any assessments that if I failed anyone who did hand in assessments I risked going over the 20% threshold. Because of this, I developed a rule of thumb that anyone who handed in all their assessments for their course would pass no matter how bad their work was. Obviously, this rule of thumb was hardly conducive to upholding academic standards. But I felt it was forced on me.

It appears my experience was not unique. 'I was party to discussions with the TEC where we tried to dissuade them from penalising universities on EFTSs funding for low pass rate courses,' a former senior administrator at AUT told us. 'They went ahead anyway and we noticed at AUT an immediate pressure to push students through.'

For a University of Auckland science professor, 'the biggest problem is the 60% required pass rate to get TEC funding' because in their subject 'we have to pass a lot of people that barely know what they are doing to hit 60%.'

A University of Canterbury academic told us about a large first-year science course he coordinates. He told us that his programme had 'a KPI [key performance indicator] requirement of an 88% pass rate,' and that 'since this is a weighted average, most of a department's pass rate' is 'determined by the pass rate of their large first year courses.' The result was 'pressures on grades' in his first-year course.

There are also more direct pressures on the grades instructors award. I interpreted the talking to I was given at the end of my first term at Victoria, for example, as an instruction to give out higher grades.

I also remember a meeting with a colleague in which she several times asked me if I would lift students above various grade boundaries (for example, from a fail grade of D to a pass grade of C-). Usually this was done with some kind of justification – appeals to mental health struggles, for example – but these were not, in my recollection, cases that involved any of the available formal channels (the aegrotat process, for instance).

In addition, I was told several times by colleagues that we had a convention of automatically lifting students whose final mark was within 1 or 1.5 percentage points of a grade boundary over

that grade boundary. So, for example, a student whose final mark was 79 would be lifted over the 80% boundary and given an A- rather than a B+, which is what her mark ordinarily would have corresponded to. And this was with grades that usually correspond to bands of only 5 percentage points each.[68]

I should emphasize once again that it is not my impression that my colleagues were doing anything unusual in this. They were, instead, reacting to the incentives that were acting upon them as a result of the system they were working within, as well as responding to the norms they saw around them in the wider university. If there was any intentionality that I could glean behind their actions, it was based on a belief that giving students slightly better grades was a kind thing to do. That belief, in turn, was, I feel sure, an aspect of my colleagues' dedication as teachers, even if I am less sure that it was right.

In any case, it seems I was not alone in being directly invited to give higher grades. A former tutor at Victoria University of Wellington told us that he was not allowed to fail any of the students to whom he had given a fail grade, and that he had to pass all students who had handed in their assessments, no matter how bad the students' work was. (In this case, it would seem that the 'rule of thumb' I applied to my own marking was an official, if not publicly-stated, policy.)

A professor in the University of Auckland Business School confirmed to us that there are sometimes 'direct interventions at the course level, when an instructor turns in a set of final results with an unacceptable number of fails.' She continued, 'I personally know of two such instances, where the instructor was explicitly told to remedy this,' although she also added that 'such direct interventions are, I think, relatively rare,' if only because with all of the other forces encouraging instructors to inflate grades, they are usually unnecessary!

The experience of a former tutor in AUT's School of Business, Economics and Law shows why. Before grading an assignment, she was handed a document containing 'marking guidelines' which she showed to us. Above a series of very basic maths questions, the document explicitly instructs tutors to give students full marks 'if there has been a proper attempt to answer the question,' adding that it does not matter 'if the solution is right.' This is not a direct instruction to give high grades, but it is hard to imagine how this kind of 'marking guidelines' would not result in higher grades being awarded.

Another feature of our model that was confirmed in our testimonies was the importance of student feedback. The University of Auckland Business School academic we quoted earlier writes:

> I have personally talked to colleagues who deliberately inflated grades in order to raise their student evaluation scores. We have every reason to do so, as the university has used low student evaluation scores to threaten employment and promotion prospects. I have heard first hand personal accounts of this. In one case a single course with an evaluation score below the "acceptable" threshold (which was set at a minimum of 70% of students agreeing that the lecturer was effective) was cited when denying a staff member confirmation as a permanent employee, and her probationary period was extended (naturally, she left nothing to chance in the next semester and easily met the threshold by inflating grades and easing student work requirements)…I was personally told by a member of our faculty staffing committee, responsible for academic promotions and job conformations, that a low student satisfaction score would threaten my reappointment. Student evaluation scores are required for all promotions, and any course with a low evaluation (even if others are good) can be used against an academic applying for promotion.

A science lecturer at the same institution painted a similar picture. 'Student reviews of courses are deemed very important by administrators,' he told us. 'But it seems like a large factor' in how students evaluate their instructors 'is simply, "was this course easy or hard?"' In other words 'when students are challenged and made to work, they give the course worse reviews' – and when instructors give students good grades, they tend to be rewarded with positive evaluations.

One claim that was repeatedly made in the testimonies we received was that New Zealand universities have in recent years made it very difficult for academics to implement robust assessment practices. Again, this echoes my own experiences. Although I often set exams in my first few years at Victoria, I had the sense that this was increasingly unusual outside of the Classics Programme. During COVID I was told I could not set exams, a situation that continued into 2022 and 2023.

Our University of Auckland science lecturer had a very similar experience over the same period:

> When I began teaching in New Zealand, exams were on paper, and done with human invigilators, as has [long] been standard... Exams are largely online now, and for several years were basically "take home exams," although there has been some move back to invigilating exams taken in computer rooms.

And our University of Auckland Business School professor described her efforts to set exams as follows:

> [There are] systematic incentives to avoid secure assessments that actually assess student competence. These vary considerably across faculties, but I can speak for my own with confidence: I have struggled for the past 5 years to have *any* meaningfully secure mode of assessment. I have not been allowed to require an in-person final exam. Supposedly secure online exams need to be justified and we have strong incentives not to go to the bother of requesting them…So like some other colleagues, I resort to holding in-person invigilated tests during term time, but we again face strong pressures not to do this.

A related issue raised by our University of Auckland science lecturer is a bias towards coursework, especially coursework of a practical nature. In his view this generally results in higher grades:

> At a large university with large introductory classes and (apparently) no funds to pay postgraduate student teaching assistants to run tutorials or set/mark homework for tutorial groups, most of the non-exam work ends up being group work done in "practicals"/labs. It is pretty easy to skate through this with high marks but low understanding. Typically the non-exam assignments have much higher marks than the exams. To ensure that passing students have a reasonable understanding of the material, some courses have a "must pass" for both practical and exam assessments. However, in my experience, in recent years the number of students who would pass on total marks but have failed to pass the exam (which requires only ~50% marks) is substantial, and it has become routine to waive the must-pass-both requirements.

Finally, our science lecturer told us that 'systems to manage cheating and plagiarism are extremely onerous for the lecturers and course administrators.' 'Outside of the most outrageous cases,' he went on, 'it is far easier to just let something slide, or use subjective marking, than to spend days on paperwork and admin to try to bring a cheating case.'

This also echoes my experiences at Victoria University of Wellington. Each year during my large first-year lecture course my tutors would bring me two or three examples of essays that had been obviously and unambiguously

plagiarized from the internet. Often these were identified by Turnitin anti-plagiarism software, which helped us to find the webpages from which reams of text had been copied and pasted.

My policy in these situations was to call the student into my office and present them with the Turnitin report on their essay. In almost every case this resulted in a quick and sometimes tearful confession from the student. I would advise them that plagiarism could have very serious consequences, they would pledge never to do it again, and that would be the end of the matter.

In a couple of cases, however, students insisted that they had not plagiarized despite the clear evidence that they had. In these cases, I would escalate the incident and try to make use of the university's formal process for plagiarism. My experience of this process was that university administrators were extremely hesitant to engage with me at all, and even more hesitant to take any action against students who had been caught red-handed plagiarizing websites and who still refused to take any responsibility for their actions.

The failure of universities to take action against plagiarism, together with the pressure on instructors to past 70–80% of students, can mean that instructors feel forced to give passing grades to students who have clearly engaged in plagiarism. If academics are forced to pass students who are guilty of plagiarism, that will obviously raise pass rates.

And raising pass rates will likely increase grades all round. Lecturers might be tempted to give honest C- level work a straight C in order to distinguish it from plagiarised work they have been forced to give the lowest passing grade of C- to. And with what would have been C- level work getting Cs, they may also give what would have been C level work a C+, and so on up the grading scale. This kind of chain reaction may play a significant role in the rising proportion of A grades we observed in Chapter 2.

Explaining the COVID spike

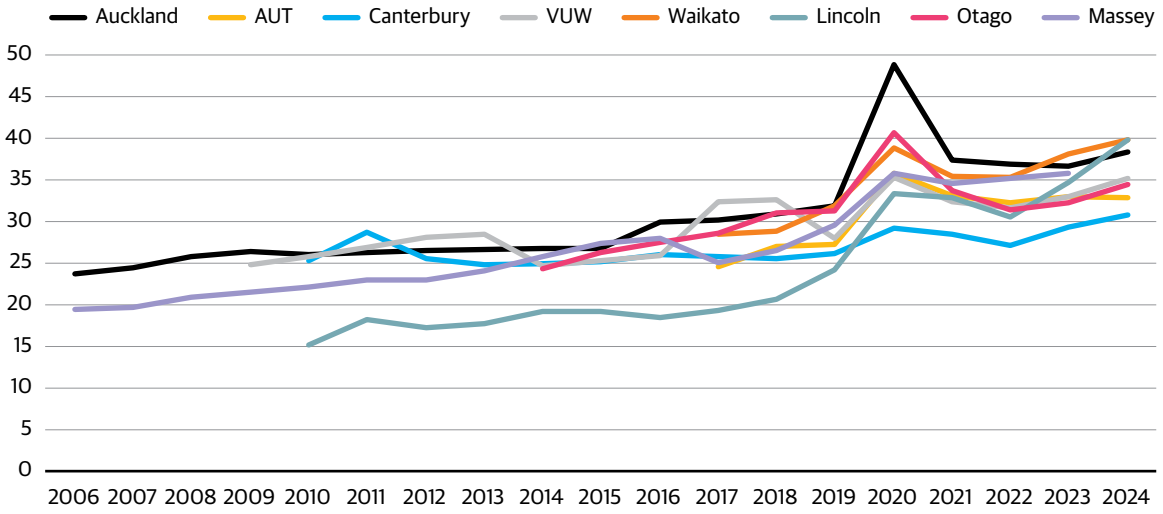
In this chapter we've set out a simple model of how grade inflation develops (our 'grade inflation game') and looked at some testimonies from academics that have given us insights into some of the pressures on grading at New Zealand universities.

The incentives and pressures we've met so far, though, have presumably long been present, and have acted on grades in a constant and steady manner. That means that though they may well explain a steady rise in grades, they can't explain the full shape of our graph of the proportion of A grades over time. Let's reprint that graph one more time (this time as Figure 4.1) to remind ourselves of what that full shape is.

What the graph shows is a gradual ramping up of the proportion of A grades until 2019; a sharp spike during COVID; and then a decline followed by a rebound in subsequent years. The incentives we have identified so far in this paper, such as the pressure of a funding model based on student numbers, have been in constant operation throughout the period of the graph, meaning that they constitute a good explanation for the gradual rise in A grades prior to COVID. But why do we observe such a dramatic spike during COVID? And why do we observe a decline and rebound in A grades after that?

Here again, there are two broad possibilities: that grade rises during COVID and thereafter reflected improved student performance, and that they were instead largely a product of grade inflation and of the incentives we have been discussing in this chapter. In Chapter 3, we saw that some of the most plausible factors that might have led to improved student performance did not explain the grade rises we described in Chapter 2. But there may have been factors that emerged only during COVID that led to improved student performance during that time.

Figure 4.1: Percentage of A-range grades at New Zealand universities, 2006–2024[69]



Source: University records via OIA requests.

For example, it might be suggested that lockdowns and other restrictions on businesses and public gatherings meant that university students had fewer distractions and more time to focus on their studies, and that this might have led to justifiably higher grades. And in fact, what effects lockdowns and other COVID measures had on university students is a complex topic with studies of various outcomes often pointing in different directions.

There does seem to be a growing consensus that COVID measures adversely affected learning outcomes among school children internationally, particularly children from less wealthy households. One study that used PISA scores from around the world found that maths scores declined by 14% of a standard deviation (i.e. by a z-score of -0.14) during the pandemic.[70] A large review of 45 studies from 15 countries estimated that students missed out on around a third of an academic year of learning.[71]

At the tertiary level, students when surveyed have tended to report adverse outcomes from COVID restrictions.[72] Fully half of the 1,500 respondents to a survey at Arizona State University reported a decrease in study hours and in academic performance during the pandemic, for example.[73] And three quarters of the 147 students who answered a survey at Victoria University of Wellington and the University of Waikato agreed that it was more difficult to focus on their studies, with just over half (54%) agreeing that their performance as a student had worsened.[74]

These are students' own impressions of their performance, and of course their impressions may be wrong. One Chinese study, though, found that students' performance on a standard test of critical thinking was worse after lockdowns than they had been before,[75] and stay-at-home orders may have had similar adverse impacts on the cognitive performance of New Zealand students.

Strictly, in order to be sure that it was online learning during the pandemic that led to downturns in academic performance, we would need to compare students who were restricted to online learning during COVID with peers in the same cohort who were able to continue to go into the classroom during the same period. Obviously, the nature of nationwide lockdowns means this was almost never a possibility.

A few studies have, though, been able to track both students who were forced to take up online learning and students in the same cohort who were able to continue with conventional face-to-face learning during the pandemic. A study of 551 students at West Point military academy found that online students' grades were 0.2 standard deviations worse than their peers' during the pandemic (i.e. they had a z-score of -0.2).[76] And a much larger study involving hundreds of thousands of students at a large public university in the US similarly found that online students' grades were worse, with a lower proportion of online students obtaining A grades than students who had face-to-face instruction.[77]

None of this is particularly promising for the idea that the spike in A grades we observed during the COVID period was a result of improvements in student performance. Moving to online instruction (which happened at all New Zealand universities during the pandemic) seems to worsen learning outcomes, and lockdowns may degrade critical thinking skills. There is no evidence that Kiwi students improved their intellectual or academic performance when COVID measures were in place, and if they did they would be global outliers.

If student performance did not improve, what we are left with is grade inflation, during the COVID period as over the last twenty or so years. But the rise in the proportion of A grades during COVID represents a particularly intense period of grade inflation, one that the factors we surveyed in the last section may not be sufficient to explain. What other factors could help explain it?

If I had to take a guess based on my own experience of the COVID period at Victoria University, I would say that this was a result of an intensification of the ethic of 'kindness' that had long helped grease the wheels of grade inflation, perhaps by allaying the consciences of academics who had been driven to give out easy grades mainly as a result of economic incentives.

I remember being encouraged several times during this period to be especially 'kind' and 'understanding' of students and their circumstances, to be aware of their mental health, and to be (even more) willing to grant extensions. If I don't remember a specific instruction to give out easier grades, I definitely came away with the impression that taking a stand against grade inflation would probably not be seen as 'kind' or 'understanding.'

A social science lecturer at VUW had similar recollections:

> From the start of the pandemic, the university administration made it quite clear that students were to be afforded every "accommodation" with their course work. This included extra time for course work submission; extra time for online responses; and being generous with our marking. The accommodations approach was to recognise that particularly Pasifika students might have greater commitments on their time in providing family care in families where parents were classed as essential workers and therefore [they might have] less ideal conditions for studying.

Conversations I have had with academics at other New Zealand universities suggest that our experiences at VUW were by no means untypical.

Some universities may have gone further than simply encouraging their academic staff to lower expectations. When I asked contacts at the University of Auckland why they thought their grades inflated so rapidly during COVID, a science professor told me that 'there was a specific University-wide grade boost of 5%' (that is, a full letter grade) 'on the instructions of the then Deputy Vice-Chancellor (Academic).'

According to this professor, 'the idea was to compensate for putative disadvantage caused by the move to online teaching,' but the measure actually constituted a simple 'increase in grades,'

not just the 'lift back to the previous distribution' that might have been more defensible.

We might finally ask why grades did not return to normal after the COVID spike – or at least to the levels seen in 2019, before the pandemic. As we saw in Chapter 2, the proportions of As awarded by New Zealand universities did fall back from the heady heights reached during COVID, but never to as low as they had been in 2019. In addition to this, grades have also had a tendency to rebound in the years since the easing of COVID measures, in the case of some universities to an even higher level than during the pandemic.

What accounts for this? On some level, the upward trends seen since the pandemic can be seen as an instance of 'normal service resumed.' Once the special circumstances of our COVID response had been removed, universities no longer had the extra pressure to inflate grades stemming from an intensified ethic of 'kindness' and from special measures such as the University of Auckland's one-off, across-the-board grade boost. But the incentives we described in our grade inflation game continued to operate, and hence grades continued to inflate.

But it may be that this final stage on our graph also illustrates the power that the COVID response had to super-charge tendencies that were already present before it, lifting trendlines onto new and higer levels. People talk about the 'new normal' that our COVID response ushered in on a variety of fronts, the increasing use of Zoom being only one example.[78]

In New Zealand, one effect of our COVID response was to increase public spending. The government justified the increases during 2020 and 2021 on the basis that it was necessary to deal with the pandemic, including by compensating businesses who were forced to close during when COVID measures were in place. But levels of state spending as a proportion of GDP have not returned to 2019 levels since, and are not projected to within this decade.[79] The declaration of emergency acted like ratchet, with spending being cranked up sharply, but without any prospect of it going down again.

The pattern of grade rises in the wake of the pandemic may be another example of this ratchet effect and of the post-COVID 'new normal.' On this model, grade rises that would have seemed inconceivable before the pandemic – to the point where almost half of the all the grades awarded at the University of Auckland were As, for example – were suddenly deemed not only appropriate, but necessary as a matter of basic decency. For a couple of years, university communities had a chance to get used to this new normal, where getting an A was as likely as getting heads on a coin toss.

As the government reduced and then did away with its COVID measures over the course of 2022, New Zealanders started to change their behaviours back to something closer to pre-COVID norms. People went out to pubs, clubs and restaurants, held in-person events (including weddings and funerals), and some even attended lectures at universities. Nature, as the saying went, was healing.

At universities, many academics must have thought that, in the absence of COVID measures, the new ethic of (extra) 'kindness' was no longer necessary. Many probably thought they could return to the old normal that had existed before COVID.

And yet, for many others in university communities, things would never be the same again. They had gotten used to the even easier grading that had been established during COVID. A taboo had been broken, and where there might once have seemed a limit, however vaguely defined, to how much academics could inflate grades, no such limit now existed. Grades could resume their apparently inexorable rise upwards.

In the next chapter, we will ask what, if anything, might stop or slow that seemingly inexorable rise.

CHAPTER 5
Deflating Grade Inflation

In my last few years as a lecturer at Victoria University of Wellington, after I sent my provisional grades to school administrators they would send me a histogram showing how many students had received each grade (from A+ downward) that year and the last time the course was taught. I was then asked to comment on the 'course outcomes and moderation process.'

I never received any prompting from anyone about what comments I should make, and I didn't feel much of a need to justify the grades in my large first-year course, which tended to have only a small number of A+s and distributions which didn't differ much year on year. When it came to my small Greek and Latin classes, though, I often felt a duty to comment on two things.

The first was how skewed the distribution tended to be towards A grades. I usually addressed this by saying that students who took Greek and Latin tended to be unusually able and dedicated. If the histogram showed that I had given more high grades than the previous year, I also commented on that, usually by simply stating that this year's class happened to be especially talented.

And that would be that. I never remember receiving any pushback from administrators, or even being asked to back up my assertions.

My impression from talking to academics at other universities in New Zealand is that my experience was fairly typical. Different universities have different measures in place (the University of Canterbury even calculates a 'difficulty index' for different courses), but these are rarely followed up. And as the results in Chapter 2 shows, whatever measures are currently in place have not done much to stem the rising tide of grade inflation.

Grade inflation, in other words, is very much still a problem. In this chapter, we look at possible solutions. Unfortunately, grade inflation is a complex problem and there's no silver bullet. There are, though, a number of different options for addressing the problem, some of which could be implemented simultaneously.

In this chapter, we present an overview of possible fixes, some combination of which might well be effective in curbing, if not completely ending, grade inflation at our universities.

We will survey three broad approaches to curbing grade inflation. The first approach intervenes in the grading system through various *moderation* techniques. The second focusses on *changing the incentives* that lead to higher grades being awarded. And the third concentrates on *changing the culture* around grading. (Of course, none of these categories are completely sealed off from one another, and many of the techniques we will look at impact the grading system, incentives, *and* the culture around grading.)

We will close the chapter with three possible paths for reform: a 'high road,' involving some of the more robust and complex techniques, a 'middle way' combining some more moderate approaches, and some 'baby steps' incorporating a few simple and easy-to-implement changes.

1. Moderation

'Moderation' may involve a number of different things. This includes having students' work marked independently by two different people, something that we will come back to in the next section.

What we will be focussing on in this section is statistical moderation techniques. These generally aim to incorporate more information about student performance into students' grades, into their grade point averages (GPAs), or into their transcripts.

To see why we might want to do this, let's consider another type of moderation, 'grading to a curve.' This usually involves forcing a certain number of results into each grade category.[80]

That can obviously help stem grade inflation. But there are also obvious problems with this approach. Some courses (and some disciplines) are harder than others, and some groups of students are better than others. Students' performances may also be distributed in different ways, with the marks in some classes all clustered together, and the marks in other classes more spread out. All this means that imposing the same distribution on every course's grades might actually undermine sound grading rather than enhancing it.

This is why we might want to incorporate more information into the grading system. We might want to indicate what the class average was, for example, or how far a given student's mark was above or below it. We might want to take account of how difficult a given discipline was, or a specific course. We might want to have a sense of how talented that year's cohort was, and even how harsh a grader the instructor was.

The basic aim of statistical moderation is to discourage instructors from giving out easy grades without over-correcting in a way that also distorts the true relationship between students' performance and the grades they are given. Providing more contextual information (a class average, say) does this by undermining instructors' ability to use letter grades to send false signals about student performance in return for student attendance.

This is why statistical moderation techniques have the potential to disrupt the grade inflation game we described in Chapter 4. If students no longer think they will be able to use their high letter grades to signal a level of performance they haven't earned, they will be less likely to gravitate towards classes with easy grading. And if instructors have a sense that giving out easy As will no longer attract more students, they should eventually start doing it less.

The more accurate signals that these techniques produce should also help remedy another problem caused by grade inflation: the 'grade compression' in the top grade categories which makes it difficult for employers to distinguish the outstanding students from the merely capable ones.

In what follows, we will provide a whistle-stop tour of the most important of these techniques, beginning with the simplest methods and working up to the most sophisticated.

We will begin with additional indicators about student performance that could be added to student transcripts. Then we will look at a method of calculating a difficulty rating for different fields. Finally, we turn to some increasingly sophisticated ways of incorporating information about courses and instructors into students' GPAs.

Additional indicators

Rank-ordering

Perhaps the simplest addition that could be made to student transcripts alongside letter grades is the rank that students achieved out of the total number of students on the course. So a student's transcript might read, for example, 'Classics 106: Ancient Civilizations: A- (27th of 252).' This could also me expressed as a percentile ('11th percentile,' say). Instructors could be allowed a limited number of ties or no ties at all to ensure that the performance of the vast majority of students could be compared. A limitation of rank-ordering is that it might suggest that students who achieved only a lowly ranking had performed

badly, whereas they might well have performed very well in an especially difficult course.

Class averages

A course's mean mark that year could also be printed alongside letter grades and percentages, as for instance 'Media Studies 301: A = 88% (average class mark: 90%).' That would immediately make clear that a high grade was achieved in a class where grades were generally high. In general, it would communicate how far a student's mark was above or below the average mark in the class. It wouldn't, however, incorporate how spread out the other grades were. If you think this isn't important, see the example in the text box.

Z-scores

A z-score is simply the difference between a figure (here, a mark) and the mean expressed in terms of a distribution's standard deviation (a measure of its spread). For example, if the mean final mark in my course is 60%, and the marks have a standard deviation of 10, then a final mark of 70 will have a z-score of +1. (Conversely, a final mark of 50 will have a z-score of -1.) Z-scores give a sense of how well a student has done compared to other students in their class in one, easily-calculated figure.

They have been recommended in both the UK and US as a way of combatting grade inflation.[81] Z-scores are slightly more complicated than class averages, but they would also make clear that Jules' performance was better than Jim's in a way that class averages would not (see the text box). So although both Jules and Jim would have the same percentage mark (85%) and class averages (70%), Jules would have a z-score of +3 and Jim of only +1. Adding these scores to their transcripts would thus add valuable information about their performance.

Difficulty ratings

Disciplinary averages

The simplest way of computing difficulty ratings for different fields is by comparing the average mark obtained in one field with the average across the whole university. So if the average grade across the whole university is 80%, the average grade in chemistry is 72% and the average grade in commerce is 84%, we could give chemistry a difficulty rating of 8 and commerce a difficulty rating of -8, say.[82] The difficulty rating for the relevant field could then be printed alongside latter grades and percentage marks on student transcripts, as for example 'Chemistry 201: A- (82%; field difficulty rating: 8).'

Average of pairwise differences[83]

One problem with computing difficulty ratings from class averages is that it doesn't take into account the fact that different students often go into different fields. If the average mark in maths courses is only 68% and the average mark in theatre is 80%, it's at least a possibility that this is not because maths is more difficult, but because students who choose to take maths are less able, or because students who choose to take theatre have a special talent for it.

The importance of the spread of marks in how we judge performance

Say Jules gets 85% in his French course and Jim gets 85% in his history course. Imagine further that the mean mark in both classes happens to be 70%. Both Jules and Jim have obtained marks that are 15 points higher than average. But the standard deviation in Jules' course is 5 (with all the final marks clustered within a narrow band), whereas in Jim's course it's 15 (with the final marks more spread out). With this in mind, Jules' performance is clearly more exceptional than Jim's.

We can get round this by looking at the marks obtained by individual students in courses in two different fields. So if Marama gets 78% in maths and 84% in English, the difference is obviously 6 percentage points. We can then calculate the average difference in marks obtained by all the students who have taken those two courses. Looking at the whole set of marks obtained by students who have taken two courses in different fields produces difficulty indices of the sort displayed in Table 5.1, which shows the difficulty indices obtained in three different US studies (with the authors' initials shown as 'G&W' and so on).[84]

The advantage of this method is that it focusses on pairs of courses which have been taken by the same individuals, meaning that the differences in the marks they obtain cannot be a product of different students choosing to go into different fields. If individuals who take chemistry and theatre receive, a mark that is 10 percentage points higher on average in the latter, that is probably not because some individuals have a special talent for fine art and gravitate towards it.

One disadvantage is that this method focusses on students who have taken courses in two different fields. But while lots of students take courses in similar subjects (maths and physics, say) relatively few students take courses in two very different fields (maths and theatre). This means that difficulty indices calculated by this method probably understate the differences in difficulty between very different fields.

Table 5.1: Grading indices for a sample of disciplines at two US colleges in the 1970s and 1980s

Field	G&W Index	S&E (Intro) Index	S&E (All) Index	E&S Index
Anthropology	-0.09	-0.05	-0.02	0.00
Art	+0.22	-0.10	+0.09	+0.06
Asian Studies	–	–	–	+0.12
Biology	-0.53	-0.27	-0.28	-0.32
Chemistry	-0.36	-0.40	-0.32	-0.35
Comparative Lit.	–	–	–	+0.31
Drama	–	–	–	+0.37
Economics	+0.18	-0.35	-0.35	-0.44
English	-0.10	+0.07	-0.01	+0.05
Engineering	–	–	–	-0.16
Ethnic Studies	+0.38	+0.42	+0.28	–
Foreign Language	+0.06	–	–	–
French	–	+0.38	+0.23	+0.08
Geology	-0.30	-0.16	+0.18	+0.02
Geography	–	–	–	-0.16
German	–	–	–	-0.07
Government	–	+0.04	-0.04	-0.19
Greek & Latin	–	–	–	+0.05
History	+0.05	+0.23	+0.01	-0.07
Mathematics	-0.07	-0.25	-0.24	-0.37
Philosophy	+0.17	+0.01	+0.02	-0.07

Source: Reproduced from Valen Johnson, *Grade Inflation, op. cit.*

GPA adjustments

The final set of methods we will look at here incorporate information about how difficult courses are into student GPAs. The idea here is that if students know that any As they obtain by taking easy courses will count less in terms of their GPA, they will be less eager to take those courses. That in turn, will mean that instructors will be less able to attract students to their courses by giving out easy As.

Iterated adjusted GPA

In this method, the GPA of each student in a course is used to calculate a GPA for that year's class. So VUW's 2037 third-year class on advanced statistical methods has 74 students with an average GPA of 88%.[85] They are clearly excellent students. The course, however, is also notoriously difficult, and when the instructor sends his grades to the school administrator at the end of the term, the mean mark is found to be 74%.

Now imagine that a clever and hard-working student called Ziming takes the course and receives a grade of 76%. Normally Ziming would expect to do better than that, and indeed her GPA is 82%. But in this class she has received a B+, meaning only 6 points for her GPA. Ziming's GPA is at A- level, so this reduces it slightly. You might say this is regrettable, since Ziming's performance in this class may have been as strong as ever. The problem is that the class is unusually difficult. But why should she be penalized for that?

This is where the iterated adjustment process could help. So imagine that when the school administrator sees that the average mark for the 2037 class is 74% and the average GPA of the students in that class is 88%, she takes account of the difference and authorizes a small upgrade in the points that are awarded towards each student's GPA. Though Ziming still gets a B+ for the course, she now also gets 6.3 points for her GPA rather than just 6.

Once this is done for all the students, though, their GPAs will have changed, meaning that the average GPA of the students in the class will have gone up. The average grade in the class will also have gone up, at least as calculated by GPA points. But the average grade in the class will have gone up by more, since fewer grades are involved in the GPA for that year's class than in the collective GPA of all the students who happen to be taking that class (because each student's overall GPA is calculated from the grades they've obtained from all the classes they've taken, not just one).

As the school administrator iterates this process again and again, the GPA for that year's class and the average GPA of the students who've taken it will converge. The three Duke researchers who proposed this method suggest that ten iterations are usually sufficient to reduce the difference between the two figures to 0 to three decimal places, at which points the iterations should stop.[86]

Regression

Methods that make use of regression analysis have been suggested by a number of researchers.[87] To get a sense of one application of regression in this context, imagine Aumai and Bill both take a biology course. Aumai gets an A- (7 in GPA terms), while Bill gets a B+ (6). The difference between them in GPA terms is 1 in Aumai's favour. But now imagine Aumai and Bill also take the same Spanish course. In this course Aumai gets a B (5) while Bill gets an A+ (9). So the difference between them on this course is 4 in Bill's favour.

Of course, Aumai and Bill will take many more courses and will overlap with many other students at their university. In order to account for this, this technique puts all of the students' scores into an enormous matrix which records students' relative performance in each course. It then asks a computer to arrive at a single figure for each student's ability that would best explain their performance as recorded in the matrix.

Item-response theory (IRT)

Item-response theory has become widely used in testing, including in the US SAT and GRE exams. The technique can estimate a student's probability of answering a question correctly based on estimates of the student's ability and of the difficulty of the question.

It has been suggested that IRT could help produce difficulty ratings for courses as well as adjusted GPAs.[88] One advantage of IRT is that unlike regression, IRT doesn't depend on pairs of students who have taken the same course. This is because IRT can work through indirect connections. So if Aumai and Bill have taken the same course, and so have Bill and Cindy, IRT can use Bill as a bridge to estimate how Aumai and Cindy compare to one another, even if they have never taken the same class.

Computers programmed to apply IRT can perform these types of operation for large numbers of classes and students. Rather than estimating each student's ability or each course's difficulty in isolation, the model finds the combination of all student abilities and all course difficulties that best explains the entire observed pattern of grades across the whole university.

Johnson's multi-rater model

A quarter of a century ago, the statistician Valen Johnson designed his own system of grade moderation.[89] Johnson's system is an item-response model that takes into account the fact that different instructors grade differently (hence 'multi-rater'). It also takes into account the fact that students might perform differently in different classes, defining student ability as the mean in a distribution of performance across various classes.

All of this means that Johnson's model upholds instructors' judgments about the relative performance of students better than the traditional GPA. With the traditional GPA, it is possible for instructors to rate Aumai higher than Cindy and Cindy ahead of Bill, but for Bill to get a higher final GPA then Cindy and Cindy to get a higher GPA than Aumai, mainly because of what classes the students choose to take. Johnson's model ensures that the final GPA instead mirrors the consensus ranking of instructors.

When Johnson tested his model with 1400 Duke undergraduates, the GPAs produced were closer to students' SATs than GPAs produced by traditional and regression approaches, and Johnson's technique was also better at predicting which of two students would get a better grade in a particular course.

As a specialized adaptation of item-response theory, Johnson's model is perhaps the most sophisticated method for integrating information about course difficulty into student GPAs.

At the same time, it suffers from the same major drawback of these approaches, which is that their sophistication makes them extremely difficult to understand. Even the way that course grades would no longer always contribute the same number of points to GPAs would require a lot of explaining to the public and to students. The fact that these methods would effectively constitute a 'black box' or opaque procedure that would often lead to lower GPAs would obviously not help matters.

This, though, is not why Johnson failed to have his tailor-made technique adopted at Duke, with the university's Arts and Sciences Council voting 19 to 14 against the proposal. On Johnson's account, the opposition to his model came mainly from social science and humanities professors.[90]

Could it be that they wanted to retain their ability to attract students to their courses by offering easy grades? If so, the episode illustrates the powerlessness of even the most thoughtful statistical moderation techniques when faced with the incentives that continue to drive grade inflation at universities. It is to these incentives that we should now turn.

2. Changing the incentives

It was one of the biggest days of my life, and I was dressed like a penguin. Black suit, white shirt, white bow tie. Adding the finishing touches, I pulled on a thin black gown and picked up a black mortar-board cap. Then I walked along the cobble-stone street towards whatever fate awaited me.

That was the first day of my final exams at Oxford. Oxford has a number of antiquated conventions around examinations, the most obvious of which is the requirement that all examinees wear the formal attire known as 'sub-fusc,' which includes a gown and an academic cap. But the Oxford exam system also has a number of conventions that might well be of interest to us in our quest to find ways of constraining grade inflation.

We can start with where I was headed that day, to a grand Jacobean revival building called 'Examination Schools' that was on High Street. For most of the year the building housed lectures, but at the end of each year it transformed into the main venue for university exams. That meant that students from all of the university's 39 constituent colleges had to make their way to this one, central building to have their academic performance tried and tested.

Oxford, like Cambridge and a few other UK universities, has a federal structure. Students live in their colleges and receive the most important part of their education (tutorials with fellows) in that environment. I studied classics with academics who taught me ancient history, philosophy and literature in my college. They gave me feedback on my essays, and would also occasionally set and mark mock exams to help me prepare for the real thing.

When it came to that real thing, though – the final exams – my tutors couldn't help me. When I sat down in one of Examinations Schools' cavernous auditoria, I wrote a candidate number, not my name, on my exam script. I did this knowing that it was part of a system in which my tutors might play a role in grading my script, but in which they would have no way of identifying it as mine.

Of course, they knew all this too. They knew that they would not be able to give me a better grade even if they happened to like me, had the impression that I worked hard during the term, or even if they wanted me to be successful so that it reflected well on them. Because of that, they had fewer incentives to give out generous grades as examiners.

They also had very little reason to give me generous grades during term. Most of them would write a percentage mark on each of my weekly essays and on my mock exams. But these were a kind of pseudo-grade whose only real purpose was to indicate the level my work was reaching. I could not have used the marks on my weekly essays in an application for a job or to a postgraduate programme.

My college tutors did have strong incentives to do a good job of preparing me for university exams, though. This was because university exams were an external and respected test of the preparedness of undergraduate students from different colleges. It would reflect poorly on a college and its tutors if their charges routinely performed badly in Examination Schools.

Between 2005 and 2024, the university published the Norrington Table, a ranking of all the colleges by their exam results, which likely helped increase competition between the colleges. Note, though, that since none of the colleges graded their own students' work, they had no option of simply handing out easier grades.

All these dynamics, I would submit, made for incentives that helped constrain grade inflation. In fact, Oxford experienced the smallest rise in the proportion of first-class degrees in England between 2010/11, when 29% of students got a first degree, and 2021/22, when 36% did.[91]

Obviously, grades have risen at Oxford as elsewhere, and it seems unlikely that this can be explained entirely by improvements in student performance. But the Oxford system might provide a good starting-point for thinking about how we might fight grade inflation by undermining some of the incentives that lead to grade inflation in the first place.

National disciplinary exams

Would it be possible to introduce something like Oxford's examination system in New Zealand – minus the very un-Kiwi formal dress?

The most obvious move would be to have each New Zealand university hold university-wide exams. But this wouldn't work. New Zealand universities aren't made up of different colleges all of which could compete for the best grades in a centrally-administered exam. There would also be too few academics in most of the programmes at our universities to ensure that graders weren't grading their own students.

This is where national-level exams might come in – something that there's a German word for. The *Staatsexamen* or 'state exams' are national-level exams that doctors, lawyers, teachers and some other professionals have to pass in order to be allowed to practice their profession in Germany. Crucially, these exams are administered not by universities but by the relevant government ministry.

As in the Oxford system, there is thus a separation between grades and instructors, but on an even larger scale. German universities compete to prepare students as well as possible for an external benchmark, just as Oxford colleges do. And, as with Oxford colleges, German universities do not have the option of simply giving out more top grades, since they aren't the ones running the exams.

Though New Zealand's universities aren't federal, they used to be part of a single University of New Zealand until 1961. That makes them ideally suited for national-level exams. New Zealand also places less emphasis than other countries on state-administered medical or legal exams that control entry to those professions, preferring to have doctors qualify through the two university medical schools (Otago and Auckland), and lawyers via the Professional Legal Studies Course (PLSC). That leaves space, not only for state exams in these subjects, but for national-level exams in other subjects as well.

We might, in fact, consider having state exams in all of the disciplines that Germany currently has them for. This would include, as well as doctors, dentists, physical therapists, lawyers, notaries, teachers, research librarians, archivists, pharmacists, food chemists and psychotherapists. This would obviously only cover a narrow range of academic disciplines, and we might want to consider also having national exams in other well-established academic fields too.

Of course there would be practical challenges. One problem is that the New Zealand university system offers individual programmes and instructors considerable leeway in what they choose to offer students. So one classics programme may choose to focus more on history and archaeology, while another might play to its strengths in the study of Greek and Latin literature and language.

Our sense is that most New Zealand university programmes nonetheless try to ensure that undergraduates are exposed to the basics of their fields. Some even require majors to take a number of different courses precisely to ensure broad coverage of the fundamentals of their disciplines. And this is what national disciplinary exams would test – the core material in any given subject, as agreed upon by subject experts.

Of course there would be arguments and horse-trading about what should be in the exam, but there would likely be quite a lot of

agreement about the main things undergraduates should be expected to know. If turf-wars within New Zealand academia became a problem, with some academics pursuing personal feuds by trying to get their rivals' favoured sub-fields excluded, a simple solution would be to have panels of international experts draw up the exams.

International experts could also be involved in grading exam scripts. All of this should lead to rigorous, 'gold-standard' exams testing the core curricula of important disciplines. Universities would then want to compete to make sure that their students performed well on these exams.

Initially, there would be a widespread sense that the grades awarded by the exams were more reliable than the grades given out by university programmes, which would still have incentives to give their own students good grades. With time, though, university programmes would also likely begin to grade more soundly, both to reduce embarrassing gaps between their marks and marks obtained by the same students in the national exams, and to give students an accurate sense of how well they would be likely to perform in those exams.

External/international moderation as part of funding formulae

Uncoupling grading and teaching (one of the virtues of national disciplinary exams) could be done more simply by having external moderation of anonymized exam scripts and/or coursework by panels of academics from other New Zealand universities. We could also have external moderation by international academics, or by a combination of international and New Zealand academics.

Of course, having all the assessments at our universities cross-marked by academics at home or abroad would be extremely costly and time-consuming. Even trying to cross-mark most or half of all the assessments that take place in undergraduate courses would be very onerous. Expert panels could instead mark samples of assessments from individual programmes, school, and/or universities.

The marks given by the external panel could then be compared to those given by the units in question. How closely each unit's own academics matched the marking of the external experts could then be fed into a formula for determining that unit's future funding. That would introduce negative incentives against grade inflation that might to some extent counteract the incentives we discussed in Chapter 4.

De-emphasizing student numbers in funding

One of the key elements in the 'grade inflation game' we described in our last chapter is that funding is dependent on student numbers.

This was one of the main outcomes of the market-oriented reforms to universities of the early 1990s.[92] Doing away with it entirely would reverse one of the central planks of the university system as it has functioned for the last three decades. It would likely be extremely difficult to pull off politically, and would involved wide-ranging disruption to the system.

This might still be worth doing if the results of the market-oriented reforms were uniformly negative. But this would be a bold judgment. Competition tends to drive up performance, and we believe that competition for students has played a key role in making New Zealand universities more open and pragmatic places, as well as inculcating a service ethos and making academics more responsive to students.

If completely removing student numbers from the determination of funding may be a non-starter, though, reducing the weighting it is currently given may not be. Until 2023 student numbers were factored into funding through the Student Achievement Component (SAC) in funding; since then this has been effected through the new DQ ('delivery + qualification') mechanism.[93]

Student numbers have traditionally had a large weighting in funding allocations alongside other mechanisms, such as the Performance Based Research Fund (PBRF), a measure of programmes' research productivity (though this too has recently been scrapped).[94] One way in which the impact of student numbers could be reduced is simply by reducing its weighting in the allocation of total funding to programmes vis-à-vis other considerations, such as research productivity.

De-emphasizing student feedback in instructor evaluations

Student feedback is now central to how academics are evaluated on their teaching, both in internal promotion applications and in applications for other academic roles.

But it has proven difficult to find any clear correlation between student feedback scores and student learning outcomes.[95] And giving out higher grades usually leads to better student evaluations, which has been a key driver of grade inflation.[96]

What students think of their instructor should clearly play some role in assessing academics' teaching. But it should be less important to teaching assessment than they are now.

Assessment of teaching by peers (that is, other academics) should become more routine. Student learning outcomes, perhaps measured by national disciplinary exams, should also be factored in.

For the time being, we would suggest that academic administrators simply give teaching evaluations far less weight in assessments of academics' teaching. This is should in itself lessen incentives on instructors to deflate grades.

Making universities liable for student loans

The British higher education consultant Peter Ainsworth has recently suggested that universities issue student loans rather than the state.[97] Under his scheme, universities would issue income-contingent loans that students would pay back as soon as they were making a large enough income.

The idea is to incentivise universities to select only those students who are prepared for university study, and to educate them in a way that makes them attractive to employers as soon as they graduate. The better a job that universities do at that, the sooner their loans will be repaid. Ainsworth notes that a similar idea – colleges being liable for student loan defaults – has recently been floated in the US in a bill introduced by Representative Virginia Foxx, Chair of the House Committee on Education and the Workforce.[98]

One disadvantage is that courses in the humanities and in the pure sciences might be disincentivised, with universities less likely to bet on graduates in these fields making good money within a few years of graduation.

But a change along these lines could have a positive impact on grade inflation, with the job market providing an external and independent check on universities' judgment of students. Universities would be aware that giving out easy As wouldn't necessarily lead to their loans being repaid any faster. What would is preparing students genuinely well for employment. And that, in turn, might lead universities to use grades more as signals to students of how good their work is.

In this country, changes to the student loan scheme administered by StudyLink would likely be controversial. There would be understandable concerns about a possible narrowing of access. The state could be asked to be a lender of last resort for students that universities were unwilling to lend to for reasons outside of those students' control. Even so, worries about universities gaining too much control over students might make this option seem less palatable in New Zealand than in the US or UK.

3. Changing the culture

In this report we've mainly focused on incentives. Alongside incentives, though, there may also be cultural issues that help support grade inflation.

Foremost among them is the ethic of 'kindness' that I mentioned in the first chapter, which dictates that instructors should be kind to individual students by bumping them up to a higher grade that they have not quite achieved. This in spite of the fact that the more instructors do this, the more problems we will have with grade inflation.

Our final category of interventions, then, concerns possible changes to the culture around grading at our institutions. All of them are ultimately concerned with transforming the ethic of 'kindness' we have just described into an ethic of fairness and accuracy.

From 'kindness' to fairness

Giving students the grades that they deserve can feel cruel. That's especially the case if we know the student well, and know that they have put in a lot of effort. And what if our giving an appropriate grade leads to a student missing out on an important opportunity?

Conversely, giving out good grades can often feel like kindness, especially when we make an exception and lift someone above what might seem like a senseless, bureaucratic dividing line between an A and a B (for example). Going though the proper channels to document an issue in a student's life (such as a mental health struggle) can often feel onerous and invasive, so isn't simply taking account of this in the student's grade the kind thing to do?

As we discussed in the previous chapter, though, what might seem like kindness to an individual student contributes to a phenomenon which has real costs for society at large. Contributing to grade inflation might thus be seen as an anti-social act comparable to depleting a common resource.

It is also arguably immoral in other ways too. Instructors who knowingly give out higher grades in their class than is the norm are effectively unilaterally claiming a right to give their students better grades for less effort, in a way that disadvantages peers who have taken classes with fairer grading.[99] This could be said to constitute a kind of arrogance – the arrogance of believing that your students matter more than other students, just because they have chosen to take your class.

Alongside changes to incentives, then, there is a pressing need to change the culture around grading at our universities. We suggest that New Zealand universities should shift from a culture of 'kindness' in grading to one of fairness.

Our concept of 'fair grading' emphasizes the moral values that animate the 'sound grading' we described in Chapter 4. Sound grading, for us, embodies two main values, accuracy and equity.

Let's start with accuracy. Grade inflation degrades grading by making it less accurate. As well as putting short-term individual needs ahead of the common good, grade inflation also exemplifies an intellectual or academic failing. Always striving to be as accurate as possible is (or should be) an academic virtue, and thus a lack of accuracy in any domain says bad things about academics. Academics should be more mindful that this applies to the grades they give out as much as to the papers that they publish.

The second value behind sound grading is equity.[100] Accurate grading is fair grading because it treats students as equals. If students are awarded worse grades just because of the courses they happen to have chosen or the instructor who happens to have taught them, that is an injustice.

It may seem like a minor injustice in any particular instance, but academics should strive to minimise this sort of injustice nevertheless. And missing out on an important opportunity

because your instructor happened to grade more fairly than other instructors can represent quite a significant injustice in the life of young person.

Our recommendation in this sub-section, then, is simply that everyone with a stake in New Zealand universities should work to shift the culture around grading from the 'kindness' that has dominated for so long to a notion of 'fair grading' grounded on the values of accuracy and equity.

The next two sections briefly sketch out a couple of mechanisms that might help us achieve this shift in the mindset.

Educating educators about grade inflation

One simple way of making sure instructors are fully aware of the temptations and the costs of grade inflation would be to integrate a module on the topic into the initiation programmes that all new academic staff have to go through.

Optional follow-up trainings could also be part of the offerings of universities' in-house centres for academic development, which deal with teaching enhancement and innovation. These trainings could also be made compulsory in the cases of academics with particularly egregious histories of handing out easy grades.

Training sessions on grade inflation could educate instructors on what grade inflation is, what its consequences are, and how it arises. They could also incorporate data on the current state of grade inflation in New Zealand, provided either by universities themselves or by a national body such as the Tertiary Education Commission.

Lecturers and professors should also educate their tutors about grade inflation, role model good moderation practices, and discourage over-generous grading on the part of tutors.

Prizes for sound grading

A final way of changing the culture of grading at our universities would be to award prizes for accurate grading. Obviously awarding prizes for the harshest grading wouldn't do, as then programmes might vie to give out the lowest grades, thus producing grading that was excessively harsh rather than excessively generous.

We could judge the accuracy of the grades given by instructors by comparing them with the grades given to the same work (or, more practically, a sample of it) by panels of international experts. Earlier in this chapter we suggested something similar, with grades awarded by external moderators (either international, domestic, or both) being used to assess instructors' grades, and the results feeding into funding formulae for programmes, schools and/or universities.

The suggestion here, though, is that non-monetary prizes be awarded purely to bestow praise and prestige on units engaging in sound grading practices. A 'Sound Grading Award' might be given to the best programme, school, and university in the country, and perhaps even to the best-performing academics. This could be done at a gala evening, where the awards would hopefully be picked up by the press. The international committees' grades could also be used to produce an annual league table of universities by the accuracy of their grading.

One risk of this is that it might backfire, with the majority of students (most of whom aren't exceptional) deciding to stay away from programmes that engage in sound grading. Since sound grading does offer some benefits to all students, though (giving them a realistic sense of their strengths and weaknesses, for example) this can hardly be considered a certainty. Indeed, just giving out a national award for sound grading might help raise its profile, making people more aware of how integral it is to true academic best practice.

Recommendations

In this chapter we have surveyed a range of possible responses to the problem of grade inflation in this country. They range from extremely sophisticated statistical approaches (such as Johnson's multi-rater item response model) and enormous policy transformations (such as doing away with funding based on student numbers) to simpler fixes (reporting class averages alongside student grades).

There is no simple, costless fix for grade inflation. As we noted in each case, all of the possibilities we surveyed above have downsides as well as upsides. What we will do in this final section is offer three paths to curbing grade inflation. Rather than moving from the simplest to the most sophisticated solutions, though, as we did earlier in the chapter, we will start with more ambitious solutions and progress to simpler fixes.

The first path focuses on complex interventions that might well substantially reduce grade inflation, but which would also be difficult to effect. The second path involves a mix of policies that would slow grade inflation to some extent without being too complicated to implement. The third path showcases simple changes that could be made virtually overnight with very little risk, but with correspondingly little impact on grade inflation.

The high road

Ambitious though the high road aims to be, we would not recommend completely doing away with student numbers as a determinant of funding. Making universities liable for student loans would also be too radical a change at this stage.

The high road should, though, involve putting in the effort necessary to introduce one of the more sophisticated statistical moderation techniques, such as Johnson's multi-rater model. An individual university or faculty could introduce Johnson's system on a trial basis alongside traditional grading. Ideally, this would be done with the enthusiastic participation of the university or faculty involved, but it would not be surprising at all if there was resistance of the sort Johnson himself faced at Duke.

The Education and Training Act gives universities the 'right to teach and assess students in the manner that they consider best promotes learning,' so the government couldn't compel universities to trial Johnson's method, but it could incentivise it by funding the trial and perhaps even offering additional financial incentives.

In addition to statistical moderation approaches, the high road should involve setting up national disciplinary exams, beginning with the subjects that have *Staatsexamen* in Germany.

The boards administering these exams (ideally composed of international experts as well as domestic academics) could also grade samples of internal assessments. This could in turn be used to publish league tables of universities by grading accuracy, and to award non-monetary prizes for sound grading at a well-publicised annual gala.

The middle way

A middle way might involve trialling one of the simpler statistical moderation techniques, such as regression or iterated adjusted GPA. These methods would likely still face resistance (not least because of the way they recalculate GPAs), but they would be easier to explain to stakeholders than item response approaches.

A middle way could also feature national disciplinary exams exclusively for law and medicine. Or it could dispense with national disciplinary exams entirely and simply have international committees grading samples of exams each year. Their grades could be used to produce ratings for the grading accuracy of academic units, and these ratings could be fed into their funding formulae.

Student numbers could also be de-emphasized in funding allocations, and teaching evaluations de-emphasized in the assessment of instructors. Finally, all university graders could be made to attend training on grade inflation, its temptations, and its dangers.

Baby steps

The simplest thing universities could do is simply start including supplementary information on student transcripts alongside letter grades. This might include students' percentage marks, students' rank in their class, class averages, and students' z-scores.

Difficulty indices for different disciplines could also be calculated and made publicly available, perhaps on university websites. If (as seems likely) academics in the easier fields resisted this, these could be calculated and published independently or by government.

In the current climate, with vested interests in our taxpayer-funded universities stubbornly holding out against reform, it might prove difficult to take even these baby steps.

But greater public awareness of the problem, and more public debate about it, might go a long way towards building a coalition for effective reform. That is something we hope that this report can contribute to.

Bibliography

Ainsworth, Peter. "Shares in Students: A New Model for University Funding" (London: Institute for Economic Affairs, 2024).

———. "Universities: what can we learn from the US?" Institute of Economic Affairs Insider blog (27 May 2025).

Alden, Lyn. *Broken Money: Why Our Financial System Is Failing Us And How We Can Make It Better* (New York: Timestamp, 2023).

Altindag, Duha, Elif Filiz and Erdal Tekin. "Is Online Education Working?" NBER Working Paper No. 29113 (2021).

Anglin, Paul and Ronald Meng. "Evidence on Grades and Grade Inflation at Ontario's Universities," *Canadian Public Policy/Analyse de Politiques* 26 (2000), 361-368.

Antoniou, Faye, Mohammed Alghamdi and Kosuke Kawai. "The Effect of School Size and Class Size on School Preparedness," *Frontiers in Psychology* (26 February 2024).

Arif, Rauf. "In the post COVID-19 world, Zoom is here to stay," *Forbes* (26 February 2021).

Aucejo, Esteban, Jacob French, Maria Ugalde Araya and Basit Zafar. "The Impact of COVID-19 on Student Experiences and Expectations: Evidence from a Survey," *Journal of Public Economics* 191 (2020).

Bachan, Ray. "Grade inflation in UK higher education." Society for Research into Higher Education (2018).

———. "The Drivers of Degree Classifications" (London: UK Standing Committee for Quality Assessment, 2018).

Baker, Simon. "Is grade inflation a worldwide trend?" *Times Higher Education* (28 June 2018).

Baltaru, Roxana-Diana. "Do Non-academic Professionals Enhance Universities' Performance? Reputation vs. Organisation," *Studies in Higher Education* 44 (2019), 1183-1196.

Betthäuser, Bastian, Anders Bach-Mortensen and Per Engzell. "Systematic Review and Meta-Analysis of the Evidence on Learning during the COVID-19 Pandemic," *Nature Human Behaviour* 7 (2023), 375-385.

Boring, Anne, Kellie Ottoboni and Philip Stark. "Student Evaluations of Teaching (Mostly) Do not Measure Teaching Effectiveness," *ScienceOpen Research* 0 (2016), 1-11.

Boston, Jonathan. "The crisis in tertiary education caused by inadequate funding," *Newsroom* (12 July 2023).

Bound, John, Breno Braga, Gaurav Khanna and Sarah Turner. "Public Universities: The Supply Side of Building a Skilled Workforce," *The Russell Sage Foundation Journal of the Social Sciences* 5 (2019), 43-66.

Brown, Phil and Nicholas Van Niel. "Alternative Class Ranks Using Z-Scores," *Assessment & Evaluation In Higher Education* 37 (2011).

Cameron, Michael, Barbara Fogarty-Perry and Gemma Piercy. "The Impacts of the COVID-19 Pandemic on Higher Education Students in New Zealand," *Journal of Open, Flexible, and Distance Learning* 26:1 (2022), 42-62.

Chapman, Madeleine. "C's get degrees: the extra tough law school grading at Victoria University," *The Spinoff* (6 September 2018).

Clayson, Dennis. "Student Evaluations of Teaching: Are They Related to What Students Learn? A Meta-Analysis and Review of the Literature," *Journal of Marketing Education* 31:1 (2008), 16-30.

Cline, Andrew. "Higher Spending, Lower Results: Why More Money Doesn't Equal Better Schools" (Concord: The Josiah Bartlett Center for Public Policy, 2025).

Côté, James and Anton Allahar. *Ivory Tower Blues: A University System in Crisis* (Toronto: University of Toronto Press, 2007).

Crampton, Eric and Bryce Wilkinson. "Charting the budget," *Newsroom* (7 June 2024).

Dawson, Philip and Thomas Corbin. "If uni marks are going up, does that mean there's a problem?" *The Conversation* (26 March 2024).

Deming, David and Christopher Walters. "The Impact of Price Caps and Spending Cuts on US Postsecondary Attainment" (Cambridge: National Bureau of Economic Research, 2017).

Denning, Jeffrey et al. "Why Have College Completion Rates Increased?" *American Economic Journal: Applied Economic*s 14:3 (2022), 1-29.

Draper, Joseph. "Grade inflation at English universities falls for first time in a decade," *The Independent* (20 July 2023).

Fennell, Andrew. "Graduate statistics for the UK," *StandOutCV* (28 April 2025).

Friedman, Milton. *Inflation, Causes and Consequences* (Bombay: Asian Publishing House, 1963).

Fukuyama, Francis. *The Origins of Political Order: From Prehuman Times to the French Revolution* (New York: Farrar, Straus and Giroux, 2011).

Gorelick, Evan. "Faculty report reveals average Yale College GPA, grade distributions by subject," *Yale News* (30 November 2023).

Gose, Ben. "Duke rejects plan to alter calculation of grade-point averages," *Chronicle of Higher Ed* (21 March 1997).

Hattie, John. "The Paradox of Reducing Class Size and Improving Learning Outcomes," *International Journal of Educational Research* 43 (2005), 387-425.

Hochschild, Jennifer. "Grade Inflation as a Tragedy of the Commons." In Meira Levinson and Jacob Fay (eds), *Dilemmas of Educational Ethics: Cases and Commentaries* (Cambridge, Massachusetts: Harvard University Press, 2016).

Hu, Kesong, Kaylene Godfrey, Qiping Ren, Shenlian Wang, Xuemei Yang and Qi Li. "The Impact of the COVID-19 Pandemic on College Students in USA: Two Years Later," *Psychiatry Research* 315:114685 (2022).

Jack, Patrick. "Nine out of 10 students get first or 2:1 at leading universities," *Times Higher Education* (6 February 2023).

———. "Grade deflation: first-class degrees back to pre-pandemic levels," *Times Higher Education* (8 August 2024).

Jackson, Kirabo, Rucker Johnson and Claudia Persico. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *Quarterly Journal of Economics* 131 (2016), 157-218.

Jakubowski, Maciej, Tomasz Gajderowicz and Harry Patrinos. "COVID-19, School Closures, and Student Learning Outcomes: New Global Evidence from PISA," *npj Science of Learning* 10:5 (2025).

Jephcote, Calvin, Emma Medland and Simon Lygo-Baker. "Grade inflation versus grade improvement: Are our students getting more intelligent?" *Assessment & Evaluation In Higher Education* 46:11 (2020), 1-25.

Johnson, Valen. "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science* 12: 4 (1997), 251-269.

———. *Grade Inflation: A Crisis in College Education* (New York: Springer, 2003).

Kaufman, Amy and Julia Colyar. "Canadian Academia and the Faculty Gender Gap" (Higher Education Quality Council of Ontario: Toronto, 2022).

Kierstead, James and Michael Johnston. *Blessing or Bloat? Non-Academic Staffing at New Zealand Universities in Comparative Perspective* (Wellington: The New Zealand Initiative, 2023).

Kofoed, Michael, Lucas Gebhart, Dallas Gilmore and Ryan Moschitto. "Zooming to Class? Experimental Evidence on College Students' Online Learning during COVID-19," *American Economic Review: Insights* 6:3 (2021), 324-340.

Lievore, Ilaria and Moris Triventi. "Do Teacher and Classroom Characteristics Affect the Way in which Girls and Boys are Graded?" *British Journal of Sociology of Education* 44 (2022), 97-122.

Lipson, Briar. "Spoiled by Choice: How NCEA Hampers Education, and What it Needs to Succeed" (Wellington: The New Zealand Initiative).

Lister, Zoe. "Grade Inflation at McGill is not a Myth," *The McGill Daily* (12 September 2022).

Lu, Xiaojing, Juanjuan Ma, Brinthaupt Thomas, Shaochun Zhao and Xuezhu Ren. "Impacts of University Lockdown during the Coronavirus Pandemic on College Students' Academic Achievement and Critical Thinking: A Longitudinal Study," *Frontiers in Psychology* 13:995784 (2022).

Mayer, Greg. "Grade Inflation at the University of Waterloo." Opportunities and New Directions Conference presentation (6 May 2009).

Mermin, Jeffrey, Garrett Mitchener and John Thacker. "Alternatives to the Grade Point Average for Ranking Students," *UMAP Journal* 19 (1998), 279-299.

Nam, Jane. "Grade inflation in college: trends and why it happens," *Best Colleges* (23 May 2024).

Nickolaus, Keith. "Grade inflation at Harvard: what does it mean for academic rigor?" Crimson Education blog (9 November 2024).

Norton, Andrew. "Is grade inflation a problem in Australian universities?" Andrew Norton blog (12 March 2020).

Nye, Barbara, Larry Hedges and Spyros Konstantopoulos. "The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment," *American Educational Research Journal* 37 (2000), 123-151.

OECD. "How Have Women's Participation and Fields of Study Choice in Higher Education Evolved over Time?" (OECD: Paris, 2020).

Office for Students. "Analysis of Degree Classifications over Time: Changes in Graduate Attainment from 2010-11 to 2021-22" (London: Office for Students, 2023).

Olsen, Alan. "The gender agenda: gender differences in Australian higher education," *The Koala: International Education News* (21 January 2024).

Olson, Mancur. *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, Massachusetts: Harvard University Press, 1965).

Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press, 1990).

Parker, Kim. "What's behind the growing gap between men and women in college completion?" *Pew Research Centre* (8 November 2021).

Patterson, Glenys. "University 'Reform' in New Zealand 1984-1990: Policies and outcomes," *Access: Contemporary Issues in Education* 10 (1991), 56-70.

Radavoi, Ciprian, Carol Quadrelli and Pauline Collins. "'The customer is always right:' why some uni teachers give higher grades than students deserve," *The Conversation* (7 July 2025).

———. "Moral Responsibility for Grade Inflation: Where Does It Lie?" *J Acad Ethics* (2025).

Razeed, Abdul and Craig Mellare. "Is Distinction the New Credit? An Empirical Study of Grade Inflation at a Higher Education Institution in Australia." Unpublished manuscript.

Rojstaczer, Stuart and Chris Healy. "Grade inflation at American colleges and universities," *Grade Inflation* website.

Stock, Paul, David Penny and Ema Alter. "Female Student Participation in New Zealand Universities: Forty Years on," *New Zealand Science Review* 72 (2015), 15-19.

Tang, Cat. "More Western students getting high marks in pandemic," *The Gazette* (8 May 2021).

Te Ara: The Encyclopaedia of New Zealand. "Tertiary Education."

Vanderbiel, Robert, Gordon Scharf and Daniel Marlow. "A Regression Approach to Fairer Grading," *Siam Review* 56:2 (2014), 337-352.

Walsh, Rebecca. "Grade boosting claim 'a reality'," *NZ Herald* (10 July 2001).

White, Daniella and Lucy Carroll. "'Distinction the new credit': Grade inflation puts uni integrity at risk," *Sydney Morning Herald* (25 March 2025).

Yang, Huanxing and Chun Seng Yip. "An Economic Theory of Grade Inflation." University of Pennsylvania Working Papers (2003).

Yeritsyan, Anna, James Mjelde and Kerry Litzenberg. "Grade inflation in higher education: evidence from the United States," *Journal of Agricultural and Applied Economics* 54 (2022), 375-393.

Young, John. "Adjusting the Cumulative GPA Using Item Response Theory," *Journal of Educational Measurement* 27:2 (1999), 175-86.

Endnotes

1 Stuart Rojstaczer and Chris Healy, "Grade inflation at American colleges and universities," https://www.gradeinflation.com.

2 Keith Nickolaus, "Grade inflation at Harvard: what does it mean for academic rigor?" *Crimson Education* blog (9 November 2024).

3 Jane Nam, "Grade inflation in college: trends and why it happens," *Best Colleges* (23 May 2024).

4 Stuart Rojstaczer and Chris Healy, "Grade inflation," *op. cit*.

5 Evan Gorelick, "Faculty report reveals average Yale College GPA, grade distributions by subject," *Yale News* (30 November 2023).

6 Simon Baker, "Is grade inflation a worldwide trend?" *Times Higher Education* (28 June 2018); Patrick Jack, "Grade deflation: first-class degrees back to pre-pandemic levels," *Times Higher Education* (8 August 2024). Both pieces draw on data from the UK Higher Education Statistics Agency (HESA).

7 Ray Bachan, "Grade inflation in UK higher education," *Society for Research into Higher Education*, https://srhe.ac.uk/arc/14/0072.pdf; Patrick Jack, "Grade inflation," *op. cit.*

8 Patrick Jack, "Nine out of 10 students get first or 2:1 at leading universities," *Times Higher Education* (6 February 2023).

9 Patrick Jack, *ibid*.

10 Simon Baker, "A worldwide trend?" *op. cit*.

11 Paul Anglin and Ronald Meng, "Evidence on Grades and Grade Inflation at Ontario's Universities," *Canadian Public Policy/Analyse de Politiques* 26 (2000), 361-368.

12 Paul Anglin and Ronald Meng, "Evidence of Grades," *ibid*., 363.

13 James Côté and Anton Allahar, *Ivory Tower Blues: A University System in Crisis* (Toronto: University of Toronto Press, 2007),197.

14 Greg Mayer, "Grade Inflation at the University of Waterloo," Opportunities and New Directions Conference presentation (6 May 2009), http://www.cte.uwaterloo.ca///research_on_teaching_and_learning/documents/Mayer401A.pdf.

15 Cat Tang, "More Western students getting high marks in pandemic," *The Gazette* (8 May 2021).

16 Zoe Lister, "Grade Inflation at McGill is not a Myth," *The McGill Daily* (12 September 2022).

17 Andrew Norton, "Is grade inflation a problem in Australian universities?" Andrew Norton blog (12 March 2020).

18 Daniella White and Lucy Carroll, "'Distinction the new credit': Grade inflation puts uni integrity at risk," *Sydney Morning Herald* (25 March 2025); Abdul Razeed and Craig Mellare, "Is Distinction the New Credit? An Empirical Study of Grade Inflation at a Higher Education Institution in Australia," unpublished manuscript. We are grateful to Craig Mellare for sharing this manuscript with us.

19 Ciprian Radavoi, Carol Quadrelli and Pauline Collins, "'The customer is always right:' why some uni teachers give higher grades than students deserve", *The Conversation* (7 July 2025), summarizing Ciprian Radavoi, Carol Quadrelli and Pauline Collins, "Moral Responsibility for Grade Inflation: Where Does It Lie?" *J Acad Ethics* (2025).

20 Rebecca Walsh, "Grade boosting claim 'a reality,'" *NZ Herald* (10 July 2001).

21 Madeleine Chapman, "C's get degrees: the extra tough law school grading at Victoria University," *The Spinoff* (6 September 2018).

22 Milton Friedman, *Inflation, Causes and Consequences* (Bombay: Asian Publishing House, 1963).

23 If we wanted to maintain the parallel between grade inflation and monetary inflation, we should also speak of 'grade inflation' when grades remain the same but student performance worsens. Grades generally have not decreased over the past few decades in English-speaking universities, though, so we confine this point to an endnote.

24 See e.g. Philip Dawson and Thomas Corbin, "If uni marks are going up, does that mean there's a problem?" *The Conversation* (26 March 2024); Calvin Jephcote, Emma Medland and Simon Lygo-Baker, "Grade inflation versus grade improvement: Are our students getting more intelligent?" *Assessment & Evaluation In Higher Education* 46:11 (2020), 1-25; and Anna Yeritsyan, James Mjelde and Kerry Litzenberg, "Grade inflation in higher education: evidence from the United States," *Journal of Agricultural and Applied Economics* 54 (2022), 375–393.

25 Ray Bachan, "The Drivers of Degree Classifications" (London: UK Standing Committee for Quality Assessment, 2018); Jeffrey Denning *et al.*, "Why Have College Completion Rates Increased?" *American Economic Journal: Applied Economics* 14:3 (2022), 1-29; Abdul Razeed and Craig Mellare, "Is Distinction the New Credit?" *op. cit.*

26 Office for Students, "Analysis of Degree Classifications over Time: Changes in Graduate Attainment from 2010-11 to 2021-22" (London: Office for Students, 2023).

27 It may be, of course, that grades are rising because students could perform at the same level as in the past but have been given easier assessments. This is still grade inflation in our definition, since the same grades (e.g. As) come to signify a lower level of actual performance, i.e. achievement.

28 Formatting often differed from one university to the next, as did what was included, with some universities breaking the results down by individual courses and levels (100-, 200- and 300-level), and others not.

29 Since we were interested in total *grades* awarded, we excluded categories such as 'DNC/DNS' (Did Not Complete/Did Not Sit) and 'Withdrawn' from our analysis, as these are not grades. In the case of the University of Auckland, for example, we included all the letter grades listed in the spreadsheet we were provided (A+ to D-) plus the 'Other fail' category, but we excluded the 'DNC/DNS' and 'Other' categories. (That is for 2006-2021. In the spreadsheets for 2022-2024 we were sent later on, we excluded 'DNC or DNS,' 'Unknown,' 'Withdrawn,' and 'Not Graded/Other.') Some of the other universities gave us grade distributions with fewer categories: AUT for example provided only numbers of letter grades A to D and a 'Total' category. These broad 'total' categories may include non-grade categories, which would make proportions of As and pass rates slightly lower in these cases. When we included all of these non-grade categories in Auckland's totals, though, the result was only slightly lower proportions of As and a similar-shaped line on the graph.

30 We use the median rather than the mean because NZ universities differ greatly in their number of students. In 2023, for example, Auckland had 35,337 EFTSs, while Lincoln had only 3,123. That means that simply dividing total numbers by eight might have given a misleading impression of the proportion of As awarded in undergraduate courses at NZ universities all told. Even the median figure has some downsides, especially because we have data for only two universities at the beginning of the period.

31 All figures are headcounts, not EFTSs.

32 This may be an artefact of the way the data was presented. The file we were sent by Waikato in 2022 contained the number of letter grades each year followed only by an 'other' column, which we counted with the fails. The file we were sent this year contained several other non-grade categories, including 'restricted pass' and 'ungraded pass,' which we counted with passes. It is possible restricted and ungraded passes were included in the 'other' column prior to 2022, thus yielding artificially low pass rates.

33 In line with our purpose of examining proportions of total grades warded, we excluded the categories 'Ungraded,' 'Did not complete,' and 'Continuing' from Massey's total grades. It could be that other universities' totals as they were provided to us included some of those categories, which might partially explain Massey's extraordinarily high figures. At the same time, including those categories, or even simply 'Continuing,' in Massey's totals leads to pass rates which are even more unusually low than the present pass rates are unusually high. This suggests that the categories we chose produced roughly equivalent figures for Massey as for the other universities.

34 Not all the universities are divided into 'faculties,' with Massey preferring 'colleges' and Otago 'divisions.' For simplicity we refer to all these entities as 'faculties.'

35 I.e. AUT's Faculty of Culture and Society, Massey's Faculty of Humanities and Social Sciences, and Waikato's Faculty of Arts, Law, Psychology and Social Science from our Arts and Humanities category; AUT's Faculty of Business, Economics and Law and Lincoln's faculty of Agriculture and Economics from our Business and Economics category; AUT's Faculty of Health and Environmental Science and Waikato's Faculty of Health, Engineering, Computing and Science from our Health category; and Lincoln's Faculty of Agriculture and Life Sciences from our Science category. We also removed all the faculties in our Creative Arts/Design/Architecture category from the graph for clarity.

36 See again Ray Bachlan, "The Drivers of Degree Classifications" *op. cit.*; Jeffrey Denning e*t al.*, "Why Have College Completion Rates Increased?" *op. cit.*; Abdul Razeed and Craig Mellare, "Is Distinction the New Credit?" *op. cit.*; and Office for Students, "Analysis of Degree Classifications," *op. cit.*

37 See again e.g. Abdul Razeed and Craig Mellare, "Is Distinction the New Credit?" *op. cit.*; and Office for Students, "Analysis of Degree Classifications," *op. cit.*

38 https://www2.nzqa.govt.nz/ncea/understanding-secondary-quals/secondary-school-stats/

39 See Briar Lipson, "Spoiled by Choice: How NCEA Hampers Education, and What it Needs to Succeed" (Wellington: The New Zealand Initiative), esp. Fig.1 on p. 13 and https://www2.nzqa.govt.nz/about-us/publications/insights-papers/ncea-six-indicators/

40 See again https://www2.nzqa.govt.nz/about-us/publications/insights-papers/ncea-six-indicators/

41 OECD, "How Have Women's Participation and Fields of Study Choice in Higher Education Evolved over Time?" (OECD: Paris, 2020), 2.

42 Kim Parker, "What's behind the growing gap between men and women in college completion?" *Pew Research Centre* (8 November 2021).

43 Alan Olsen, "The gender agenda: gender differences in Australian higher education," *The Koala: International Education News* (21 January 2024).

44 Amy Kaufman and Julia Colyar, "Canadian Academia and the Faculty Gender Gap" (Higher Education Quality Council of Ontario: Toronto, 2022); Andrew Fennell, "Graduate statistics for the UK," *StandOutCV* (28 April 2025).

45 https://www.statista.com/statistics/677011/uk-degree-results-by-gender

46 See again Ray Bachlan, "The Drivers of Degree Classifications" *op. cit.*; Jeffrey Denning *et al.*, "Why Have College Completion Rates Increased?" *op. cit.*; and Abdul Razeed and Craig Mellare, "Is Distinction the New Credit?" o*p. cit*.

47 It is possible that the higher grades that female students receive is partly a result of bias. See e.g. Ilaria Lievore and Moris Triventi, "Do Teacher and Classroom Characteristics Affect the Way in which Girls and Boys are Graded?" *British Journal of Sociology of Education* 44 (2022), 97–122, which finds that instructors in Italian secondary schools give higher grades to girls than to boys with 'identical subject-specific competence.'

48 Paul Stock, David Penny and Ema Alter, "Female Student Participation in New Zealand Universities: Forty Years on," *New Zealand Science Review* 72 (2015), 15-19. The numbers in both Figure 3.2 and 3.3 apply to domestic students only, since we could not find male/female breakdowns for international students on *Education Counts*. The percentage of international students at New Zealand universities has varied between 9% and 15% over the past decade. Given this and the fact that higher female participation in tertiary education is a worldwide trend, we do not think that including female international students would have significantly altered our figures.

49 https://www.educationcounts.govt.nz/statistics/tertiary-participation > Provider-based enrolments > ENR 4: Domestic students by ethnic group, age group and gender 2015-2024.

50 Much of the research on spending and educational outcomes pertains to secondary schools. More funding leads to better outcomes: Kirabo Jackson, Rucker Johnson and Claudia Persico, "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *Quarterly Journal of Economics* 131 (2016), 157–218. Not necessarily: Andrew Cline, "Higher Spending, Lower Results: Why More Money Doesn't Equal Better Schools" (Concord: The Josiah Bartlett Center for Public Policy, 2025).

51 John Bound, Breno Braga, Gaurav Khanna and Sarah Turner, "Public Universities: The Supply Side of Building a Skilled Workforce," *The Russell Sage Foundation Journal of the Social Sciences* 5 (2019), 43-66; David Deming and Christopher Walters, "The Impact of Price Caps and Spending Cuts on US Postsecondary Attainment" (Cambridge: National Bureau of Economic Research, 2017).

52 https://www.tec.govt.nz/funding/funding-and-performance/performance/financial > Tertiary education institution financial performance by year; https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG > Download EXCEL

53 University students: https://www.educationcounts.govt.nz/statistics/tertiary-participation > Provider-based equivalent full-time students (EFTS) > EFT.35: Domestic and international equivalent full-time student units (EFTS) by subsector, provider and qualification type/NZQF level 2015-2024.

54 Jonathan Boston, "The crisis in tertiary education caused by inadequate funding," *Newsroom* (12 July 2023).

55 For just one example of a study that finds that smaller classes do lead to better outcomes, see Barbara Nye, Larry Hedges, and Spyros Konstantopoulos, "The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment," *American Educational Research Journal* 37 (2000), 123–151. For studies suggesting that 'there is little evidence for major effects on student learning' from decreased class sizes, see John Hattie, "The Paradox of Reducing Class Size and Improving Learning Outcomes," *International Journal of Educational Research* 43 (2005), 387-425.

56 Faye Antoniou, Mohammed Alghamdi and Kosuke Kawai, "The Effect of School Size and Class Size on School Preparedness," *Frontiers in Psychology* (26 February 2024).

57 Roxana-Diana Baltaru, "Do Non-academic Professionals Enhance Universities' Performance? Reputation vs. Organisation," *Studies in Higher Education* 44 (2019), 1183-1196.

58 Unlike in our earlier report (see James Kierstead and Michael Johnston, *Blessing or Bloat? Non-Academic Staffing at New Zealand Universities in Comparative Perspective*, Wellington: The New Zealand Initiative, 2023, 11, Box 3) we did not include the two small categories of 'research-only' and 'research-support' staff in our figures here, on the grounds that staff dedicated entirely (or virtually entirely) to research are unlikely to boost student performance. Instead we limited ourselves to the ministry's categories of 'academic' and 'other' staff.

59 Tertiary students: https://www.educationcounts.govt.nz/statistics/tertiary-participation > Provider-based equivalent full-time students (EFTS) > EFT.1: Domestic and international equivalent full-time student units (EFTS) in formal tertiary education 1994-2024. University students: *ibid*. EFT.35: Domestic and international equivalent full-time student units (EFTS) by subsector, provider and qualification type/NZQF level 2015-2024. Academic and non-academic staff numbers are from 'University staff 2002-2022,' a spreadsheet kindly provided to us by Nicola Marshall of the Ministry of Education.

60 See e.g. Francis Fukuyama, *The Origins of Political Order: From Prehuman Times to the French Revolution* (New York: Farrar, Straus and Giroux, 2011), 68 and Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press, 1990).

61 For a classic account, see Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, Massachusetts: Harvard University Press, 1965).

62 Huanxing Yang and Chun Seng Yip, "An Economic Theory of Grade Inflation," University of Pennsylvania Working Papers (2003) present a formal model in which universities choose to give good grades to poor students so that their graduates get good jobs and universities' prestige is boosted. Their analysis shares our basic collective action framework (with grade inflation being costly to society in the long term) but focuses more on the job market than on incentives within universities. Jennifer Hochschild, "Grade Inflation as Tragedy of the Commons," in Meira Levinson and Jacob Fay (eds), *Dilemmas of Educational Ethics: Cases and Commentaries* (Cambridge, Massachusetts: Harvard University Press, 2016) also shares our basic intuition that grade inflation is like the tragedy of the commons, although without diving too deeply into the specific incentives involved in this case.

63 See esp. Valen Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer, 2003), 47-82, and the studies he summarises there.

64 See esp. Valen Johnson, *Grade Inflation, op. cit.,* 167-194, and the studies he summarises there.

65 For 'sound money' see e.g. Lyn Alden, *Broken Money: Why Our Financial System Is Failing Us And How We Can Make It Better* (New York: Timestamp, 2023).

66 See e.g. "Tertiary Education," in *Te Ara: The Encyclopaedia of New Zealand;* Glenys Patterson, "University 'Reform' in New Zealand 1984-1990: Policies and outcomes," *Access: Contemporary Issues in Education* 10 (1991), 56-70.

67 This is my recollection from VUW, and I've talked to several other New Zealand academics who have similar stories from recent years (though they say they were told not to fail more than 15% of their students, not 20%). I have also seen a document produced by a school of science at the University of Auckland in which academics are told they will need to explain pass rates of less than 70% in their courses. Though most of my informants told me this reflected a TEC requirement, I have not been able to locate any documentation of this on the TEC website.

68 The exceptions are D (0–39%), D (40–49%) and A+ (90–100%). See again the text box on grading conventions in Chapter 1.

69 All figures are headcounts, not EFTSs.

70 Maciej Jakubowski, Tomasz Gajderowicz and Harry Patrinos, "COVID-19, School Closures, and Student Learning Outcomes: New Global Evidence from PISA," *npj Science of Learning* 10:5 (2025).

71 Bastian Betthäuser, Anders Bach-Mortensen and Per Engzell, "Systematic Review and Meta-Analysis of the Evidence on Learning during the COVID-19 Pandemic," *Nature Human Behaviour* 7 (2023), 375–385.

72 In addition to the study in the following notes, see also Kesong Hu, Kaylene Godfrey, Qiping Ren, Shenlian Wang, Xuemei Yang and Qi Li, "The Impact of the COVID-19 Pandemic on College Students in USA: Two Years Later," *Psychiatry Research* 315:114685 (2022).

73 Esteban Aucejo, Jacob French, Maria Ugalde Araya, Basit Zafar, "The Impact of COVID-19 on Student Experiences and Expectations: Evidence from a Survey," *Journal of Public Economics* 191 (2020).

74 Michael Cameron, Barbara Fogarty-Perry and Gemma Piercy, "The Impacts of the COVID-19 Pandemic on Higher Education Students in New Zealand," *Journal of Open, Flexible, and Distance Learning* 26:1 (2022), 42-62.

75 Xiaojing Lu, Juanjuan Ma, Brinthaupt Thomas, Shaochun Zhao and Xuezhu Ren, "Impacts of University Lockdown during the Coronavirus Pandemic on College Students' Academic Achievement and Critical Thinking: A Longitudinal Study," *Frontiers in Psychology* 13:995784 (2022).

76 Michael Kofoed, Lucas Gebhart, Dallas Gilmore and Ryan Moschitto, "Zooming to Class? Experimental Evidence on College Students' Online Learning during COVID-19," *American Economic Review: Insights* 6 (3): 324–40.

77 Altindag, Duha, Elif Filiz and Erdal Tekin, "Is Online Education Working?" NBER Working Paper No. 29113 (2021).

78 See e.g. Rauf Arif, "In the post COVID-19 world, Zoom is here to stay," *Forbes*, 26 February 2021.

79 Eric Crampton and Bryce Wilkinson, "Charting the budget," *Newsroom*, 7 June 2024.

80 Since distributions of more than about thirty marks automatically begin to approximate a normal distribution, 'grading to a curve' doesn't usually involve forcing marks into a normal distribution, and is thus something of a misnomer.

81 Phil Brown and Nicholas Van Niel "Alternative Class Ranks Using Z-Scores," *Assessment & Evaluation In Higher Education* 37 (2011); http://replay.tlc.aston.ac.uk/wp-content/uploads/2018/12/A-new-way-of-using-z-scores-for-examination-marks-with-unfavourable-means-and-standard-deviations-Dunne-and-Eperjesi-1.pdf. Jeffrey Mermin, Garrett Mitchener and John Thacker, "Alternatives to the Grade Point Average for Ranking Students," *UMAP Journal* 19 (1998), 279-299 also suggest replacing letter grades with 'the number of standard deviations above or below the course mean' – z-scores, effectively, though they don't use that term.

82 This seems to be how the difficulty index at Canterbury worked. We are grateful to Eric Crampton for sharing his memories of this system with us.

83 On this see esp. Valen Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer, 2003), 197-209.

84 For details of these studies see again Valen Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer, 2003), 197-209.

85 In what follows, we will be using GPAs (really just grade averages) in percentage form rather than in the 9-point form that is more common in this country. This is to aid the clarity of the exposition.

86 Jeffrey Mermin, Garrett Mitchener and John Thacker, "Alternatives to the Grade Point Average," *op. cit*.

87 See e.g. Jeffrey Mermin *et al.,* "Alternatives to the Grade Point Average," *op. cit*.; Valen Johnson, *Grade Inflation, op. cit.*, 211-218; Robert Vanderbiel, Gordon Scharf and Daniel Marlow, "A Regression Approach to Fairer Grading," *Siam Review* 56:2 (2014), 337-352.

88 See e.g. John Young, "Adjusting the Cumulative GPA Using Item Response Theory," *Journal of Educational Measurement* 27:2 (1999), 175-86. Johnson's version of IRT, described immediately below, was actually trialled at Duke.

89 Valen Johnson, "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science* 12: 4 (1997), 251-269.

90 Ben Gose, "Duke rejects plan to alter calculation of grade-point averages," *Chronicle of Higher Ed,* 21 March 1997.

91 Joseph Draper, "Grade inflation at English universities falls for first time in a decade," *The Independent* (20 July 2023).

92 See again "Tertiary Education," in *Te Ara: The Encyclopaedia of New Zealand* and Glenys Patterson, "University 'Reform' in New Zealand 1984-1990: Policies and outcomes," *Access: Contemporary Issues in Education* 10 (1991), 56-70.

93 See e.g. https://www.tec.govt.nz/vocational-education/vocational-education/about-vocational-education/vocational-education-news/vocational-education-now-february-2023/dq-replaces-sac

94 See e.g. https://www.tec.govt.nz/funding/funding-and-performance/funding/fund-finder/pbrf/quality-evaluation-2026

95 Anne Boring, Kellie Ottoboni and Philip Stark, "Student Evaluations of Teaching (Mostly) Do not Measure Teaching Effectiveness," *ScienceOpen Research* 0 (2016), 1-11; Dennis Clayson, "Student Evaluations of Teaching: Are They Related to What Students Learn? A Meta-Analysis and Review of the Literature," *Journal of Marketing Education* 31:1 (2008), 16-30.

96 See again Valen Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer, 2003), 47-82, and the studies he summarises there.

97 Peter Ainsworth, "Shares in Students: A New Model for University Funding" (London: Institute for Economic Affairs, 2024).

98 Peter Ainsworth, "Universities: what can we learn from the US?" Institute of Economic Affairs Insider blog, 27 May 2025.

99 Valen Johnson, *Grade Inflation, op. cit.*, 222-223.

100 The term is also used by Valen Johnson, with 'Grading Equity' the title of Chapter 7 of his book *Grade Inflation (op. cit.*)

'Grade inflation' – when grades rise in a way that cannot be explained by improvements in student performance – has long been a major concern in university sectors across the English-speaking world. Is it a problem in New Zealand universities, too?

In this report, we show that it is. Drawing on data from each of New Zealand's eight universities, we find that the proportion of A grades and pass rates has both risen over the past couple of decades, with an especially dramatic spike during the COVID years. And we show that these increases are unlikely to be explained by students being more advanced when they arrive at university, by greater proportions of female students (who tend to get better grades), by increases in funding, or by increases in the ratio of staff to students.

Why has this happened? We suggest that this situation has developed as a result of a 'grade inflation game' that our academics have been largely forced to play. This game presents instructors with plenty of motivation to give out higher grades and very few reasons not to. We also draw on testimonies from academics that give us a sense of the pressure they feel to give out ever-higher grades.

International experience suggests that grade inflation is a tough problem to solve. But we lay out three possible responses: applying statistical moderation techniques, changing the incentives that lead to grade inflation, and changing the culture around grading at our universities. With a concerted push for reform, some combination of these approaches may well be able to start deflating the grade inflation that we document.

$25.00

ISSN
2624-0092 (print)
2624-0106 (online)

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

The New Zealand Initiative
PO Box 10147
Wellington 6143