Recognition incentives and student outcomes: Evidence from a quasi-experiment

Leigh Neethling[1]

August 2019

Incomplete draft version: Do not quote!

ABSTRACT

Many universities and colleges worldwide implement strategies that encourage and recognize student performance. This paper examines the impact of academic reward strategies that recognize students' academic performance at a large university in South Africa. Using a regression discontinuity approach to estimate the effects of the Dean's Merit List on a broad range of student outcomes, results suggest that some students who are treated with the Dean's Merit List exhibit both short and long-term declines in performance. Interestingly findings suggest that these declines intensify over time while varying in impact across faculties. These results affirm that the nature and timing of recognition policies set by academic administrators cannot be assumed to have the desired impact on heterogeneous student bodies, especially in a developing country context.

Keywords: Higher education; analysis of education; academic recognition, regression discontinuity design.
JEL Classification Numbers: I23, I21, D04

[1] School of Economics, University of Cape Town, South Africa. Email: leigh.neethling@uct.ac.za

1 Introduction

Many universities implement policies of reward or reprimand as a means of ensuring students meet academic continuation requirements from year to year. These policies may take the form of static recognition (warning) during or at the end of the academic year, or dynamic warning systems that alert students to a below-expected level of performance throughout the academic year. University administrators believe there is an effort response to these systems irrespective of the system, and by implication a performance response, without knowing the true impact of these policies.

At the institution under consideration, students receive feedback about their annual academic performance at the end of every academic year. This feedback usually takes the form of an updated academic status describing the student's annual academic standing based the number of courses passed and the student's grade point average (GPA). This information appears on students' academic transcripts for each year that they are registered.

The Dean's Merit List (DML) is a key measure used by higher education (HE) institutions worldwide to recognize and incentivize the academic performance of high-achieving students. There are two potential explanations for its implementation. The first is a reward to high-achieving students. For students whose performance exceed a given cutoff, a recognition of academic excellence is added to their academic transcript and noted through the receipt of a letter to the awardee. This list is updated annually. Students may appear on the list in one year but not meet the requirements in subsequent years. They may reappear in later years, depending on course and overall performance. The second reason is an aspirational goal for students who are just below the cutoff. If students just miss the cutoff but are informed by classmates of the recognition that occurs just above the threshold, students may be incentivized to increase their effort in subsequent years to gain access to the group of students recognized by the university.

Despite the popularity of the DML system around the world, the literature remains sparse on its effects on student performance and academic outcomes. The reasons for the lack of evaluation are two-fold. First, the lack of data in HE is widespread. Many HE institutions do not have sufficiently detailed records, mostly due to poor record keeping. Within this, data records are often incomplete with missing data a key challenge to evaluating policies and student outcomes. Second, many studies are largely descriptive or correlational in nature, failing to account for group differences or selection into award of the policy. Fortunately, with the emergence of techniques that assist to identify causal effects, the volume of research in this area has started to increase.

This paper uses a regression discontinuity (RD) design where students who earn GPA's just above or just below a given threshold are compared. Therefore, a good counterfactual is presented for a student just below the threshold and did not receive academic recognition with a student just above the threshold who has received academic recognition. Given the level at which the reward is located, both groups of students are expected to perform relatively well, and the RD design allows the evaluation of the additional impact of the policy, in addition to the already good performance demonstrated by the students (Fletcher and Tokmouline, 2010).

This paper finds significant negative effects for short-term performance, but that there is considerable heterogeneity across students for medium- and long-term effects. Humanities, male, 3 year-degree and low entry score students respond most negatively to the policy. An important finding is that these effects do not fade out over time, but rather increase in intensity between the second year of enrolment and the graduation year. For longer term outcomes such as graduation there are very small but statistically significant negative effects. Overall, there are no consistent patterns across observable characteristics. Given that none of the heterogeneous effects are consistent across faculties, there is evidence that application of strict cutoffs or absolute scores as incentives are limited in scope and impact in the context in which they are examined.

Several contributions to the literature are made through this paper. First, this paper contributes to the literature on gender and racial differences in response to educational incentives and recognition. In countries where student performance is much less understood, even more so at the post-school level, this model provides some insight into causal mechanisms that might be of (dis)incentivizing motivation to students to increase their academic performance. This information is invaluable to the academic administrator who may be uncertain about performance thresholds for short-term recognition. This paper also contributes to the effect of reward or recognition programs on developing country HE students, and specifically South African students. Wright (2018) found that incentive policies have positive and significant effects on student performance in a developing country context. However, most studies tend to focus on the developed country contexts, with little evidence available for developing country contexts where education completion rates are much lower.

The rest of the paper is organized as follows. Section 2 briefly discusses some background literature. Section 3 describes the data and institutional setting, including a detailed discussion about the Dean's Merit list programme. Section 4 describes the empirical strategy. Section 5 presents the results and section 6 concludes.

2 Background/Literature

There is a large literature examining university policies such as academic probation (Fletcher and Tokmouline, 2010), affirmative action (Massey and Mooney, 2007), financial aid (Goldrick-Rab, Douglas, Kelchen and Benson, 2012), merit aid (Leeds and DesJardins, 2015) and student advising (Bettinger and Baker, 2014) on student performance and outcomes. As much as the literature on the above policies continue to grow, the literature on policies that recognize annual academic performance is less developed, largely due to a lack of data or the manner in which the policy has been implemented.

Where research examining the effects of student reward or recognition does exist, it is mostly correlational comparing student outcomes for those who've received academic reward and those who have not (Crandall and McGhee, 1968). Comparing two groups whose observed and unobserved characteristics vary significantly leads to biased results as the two groups are not directly comparable along any given dimension.

Seaver and Quarton (1973) conducted one of the earliest causal studies on the effect of the DML on subsequent academic performance. Using a RD design, the authors find that students achieve higher than expected GPA's in the subsequent term relative to non-DML students. This effect holds true for annual GPA and cumulative GPA in the semester immediately following the award of recognition, as well as the term after that. Seaver and Quarton (1973) therefore find both short-term and long-term effects of the DML award on student performance. Unfortunately, the study does not investigate any heterogeneity of outcomes.

Wright (2018) presents another comprehensive analysis of the impact of recognition and probationary policies on student performance. Using data from a large Jamaican university the author finds that the outcomes are sensitive to the design and intensity of the programmes' implementation. Specifically, academic recognition policy effects are heterogeneous across the student body, with students in social sciences responding differently compared to students in the natural sciences. The author finds statistically significant, strong short- and long-term responses to the DML policy in the social sciences for almost all student outcomes but does not find many significant effects in the medical, pure or applied sciences. To explain these differences, the author evaluated students' course selection. Findings suggest that this significant result is driven by both strategic course taking and an increase in effort, as the improved results could lead to tangible benefits such as access to financial assistance. The author does not show if there are any gender effects.

3 Data and Institutional Setting

The data for this paper comes from the administrative records at a large university in South Africa. For each year that a student is enrolled we have data on their demographic characteristics, programme and degree registration, individual course enrolment, academic performance, financial aid and residence status. The analysis is restricted to the entering cohorts of 2006 to 2008 to exploit the uniformity in entry requirements for that time period thereby allowing us to track the students through to their exit from the system. The Health Sciences faculty is specifically excluded as the entry requirements are significantly different to other faculties. Students who completed foreign schooling and those whom there are missing variables of interest are also excluded. The sample is then restricted to those who complete their degrees in n+2 years, representative of national norms in South Africa. Lastly, the sample is limited to those within the optimal bandwidth of the DML requirement. This effectively drops any students who failed a course in the year in which the DML is awarded and those students who missed the DML requirement by significant margins as these students would not be eligible for the DML and more than likely differ significantly on observed characteristics from those who do meet the requirements for the award of the DML.

First time entering students (FU) are exposed to a detailed orientation programme before commencing their formal studies. This orientation programme usually includes the provision of important university documentation such as handbooks and the actual registration process. Students apply to and are accepted into degree programmes that are typically very structured. Degrees are least structured in the Humanities faculty where students may choose their majors and subjects with significant choice while meeting less formal programme restrictions within the degree structure. Most students register for 3-year degrees with varying course load requirements depending on the faculty of registration. The Commerce faculty has the highest course requirement for the award of the degree while the Science faculty has the lowest course requirements. The Humanities faculty is somewhere between Commerce and Science, depending on the programme of choice. Other popular degrees are typically professional 4-year degrees, usually including a fourth year of study that may be accessed as an honours degree for students completing 3-year degree versions. For many degrees offering a 4th year of study, access and progress from 3rd year to 4th year is not automatic, with minimum requirements to be met before students are admitted. This means that for many disciplines, we may observe a jump in performance for 3rd year, at least for those students attempting to gain entry to 4th year programmes.

Students' GPA is recorded on their academic transcript for each year of registration. In South Africa, grades are awarded on a scale of 0-100. The student's annual non-cumulative GPA is calculated as the weighted average of the final grades for a given academic year of study weighted by the credit weighting of the courses registered for in a particular year. Credit weightings differ by faculty due to faculties being given the autonomy to set their own credit requirements but normally imply the same required amount of hours by level of study.

At the university under consideration in this paper, students with an unrounded weighted average non-cumulative annual GPA of 70% and above are recognized for their academically excellent performance with the DML award added to their academic record. The key criteria to be recognized with the DML is completing a full course load unless students are in their final year and do not require a full load, and completing all courses during the standard semester. Students receive notification from Faculty offices via official email correspondence. Because many courses span the entire year, students' academic records are usually evaluated at the end of the academic year and not on a semester basis as may be popular internationally.

4 Empirical Strategy

4.1 Validity of the RD design

Imbens and Lemieux (2008) provide a detailed discussion on the practical implementation of the RD approach for causal identification, arguing for three assumptions to be met for any RD study to be valid. The first condition that must be satisfied is that on either side of the threshold, there is a discontinuous change in the allocation of the assignment of treatment. To qualify for the DML recognition a student must satisfy the raw, unrounded GPA requirement. Upon further evaluation of the data, it has been confirmed that the GPA cutoff at 70% is the only criteria to the awarded the DML. The second criteria to be satisfied is that of local randomization (Lee and Lemieux, 2009). Put simply, any differences in observable characteristics within the neighborhood of the discontinuity should be statistically insignificant. To evaluate this assumption the dependent variable in equation 1 is replaced with the observable characteristics of the sample to test for any significant differences above and below the threshold. The third assumption to be met is that there should not be manipulation of the running variable in the region of the discontinuity, in our instance, GPA around the award grade.

Given the "sharp" nature of the discontinuity, the following equation may be used to estimate the impact of the DML award after the first year on student outcomes:

$$\mathrm{Y_i} = \theta\ \mathrm{GPA_i^{year1}} + \delta 1\left(\mathrm{GPA_i^{year1}} > \mathrm{GPAMIN}\right) + u_i \tag{1}$$

where Y_i is a outcome for student i, $\theta(GPA_i^{year1})$ is a continuous function of the students' non-cumulative first year GPA, $1\left(\mathrm{GPA_i^{year1}} > \mathrm{GPAMIN}\right)$ is an indicator variable equal to one if the student's non-cumulative annual GPA is above the award grade and u_i is a random error term. The coefficient of interest is δ, the estimated impact of being recognized with a DML award.

The issue of bandwidth remains pertinent for this type of analysis. Upon implementation of the methodology suggested by Cattaneo, Jansson and Ma (CJM) (2018), the proposed bandwidths are very wide relative to the final bandwidths chosen. The main reasons for the final chosen bandwidths to be much narrower than the bandwidths derived from the CJM methodology are twofold. First, every indication is that the student characteristics vary significantly within this wide bandwidth, making it very likely that the second RD assumption of similar observable characteristics will be violated. Second, the upper limit crosses over into a new reward category that is related to the cumulative performance of students which is recognized at the point of graduation. This will create confounding effects that will be difficult to disentangle from the DML effect. Evidence suggests that characteristically these students differ to students with lower GPA's, making it difficult to justify the very wide bandwidths suggested by the CJM methodology.

Significant consideration was given to the choice of bandwidth. Five bandwidth options were ultimately chosen to conduct the analysis. From the manipulation tests the optimal bandwidth for usage in this chapter is derived based on the Cattaneo, Jannson and Ma[2] (2018) methodology. The CJM method indicates a lower bandwidth of 5.179% below the cutoff and 6.643% as the upper bandwidth above the cutoff value of 70%. In choosing the final bandwidths, the suggested bandwidths derived from the CJM method at $h_L = 5.179$ and $h_R = 6.643$ were retained to make the analysis more robust. As a further robustness check, estimates for bandwidths of 4.8, 4, 3 and 2 above and below the threshold are shown to demonstrate that the results are not very sensitive to the choice of bandwidth.

Further extensions present the results for the long-term impact of a DML recognition in first year. This is because a DML might be awarded in first year only with subsequent non-cumulative annual GPA falling below the cutoff of 70% for DML recognition. In addition to the long-term

[2] Cattaneo, Jannson and Ma is shortened to CJM in the text.

effects of the DML award, we are also interested in evaluating the impact of *ever* being awarded a DML.

A key consideration in any RD study is manipulation of the running variable of interest which could result in significant discontinuities at the cutoff or threshold of interest. In the context of this study, this might occur if students are able to manipulate their GPA through strategic course-taking behaviour or if they can forecast their GPA with some certainty (Casey, et. al, 2018). If students are able to manipulate their GPA then it is likely that we will observe heaping or clumping in the running variable at the threshold (Lee and Lemeiux, 2009). The resulting impact is bias of the estimated effects of the policy on student outcomes due to the extreme discontinuity at the threshold. However, given that the first-year curriculum in most faculties is fairly fixed, required courses for a chosen major or specialisation are pre-determined and that the largest contribution to final course grades is performance in the final course examination, there appears to be little opportunity for students to strategically choose courses to maximize their GPA. A further reason to support the inability of students to manipulate their GPA stems from the examinations practice at the university. Students do not use their names on examination scripts, instead students use their assigned student number on all submitted examination materials. This means that should instructors mark examinations, they will not be able to easily identify specific students and thereby award grades not reflecting the actual examination performance.

Casey, et. al. (2018) argue that if students can manipulate their GPA through strategic course selection, this will be observable in the density of GPA's around the threshold. If students are able to engage in strategic course taking, the distribution of GPA's would be discontinuous just below and just above the threshold, with very few observations below the threshold and very many observations just above. Interpreting this discontinuous allocation of GPA's, it would be simple to assume that students just above the threshold are significantly more motivated compared to students just below the threshold. However, the inability to exert significant influence over the choice of subjects prevents us from falling into this trap.

4.2 Satisfying the RD assumptions

A very detailed analysis was performed to investigate the extent to which GPA manipulation exists in the data and the extent to which it might bias results. Manipulation tests are conducted to check for evidence of a discontinuity in the density of the running variable, in this instance GPA at the cutoff of the DML award. If results show that there is a statistically significant discontinuity in annual GPA at the cutoff it could imply GPA manipulation on the part of students in the form of non-random selection or self-selection into courses, and therefore into the control and treatment groups in the analysis (Cattaneo, Jansson and Ma, 2018).

To test for manipulation of annual GPA, the procedure outlined by Cattaneo, Jansson and Ma (2018) is followed. The test for manipulation involves the estimation of the discontinuity in the density function of the running variable (McCrary, 2008). The results of the manipulation tests are based on data-driven optimal bandwidths using local polynomial density estimation for RD analysis. Figure 1 shows the local-polynomial density function plot including confidence intervals to visually and statistically examine for evidence of discontinuity or heaping for the annual GPA variable. At the DML cutoff of 70%, there is a small decrease in the density of GPA at the threshold. The estimated discontinuity of the density is not statistically significant, even though we note a big jump from an annual GPA of 69.99 to exactly 70 across the three cohorts under consideration. All manipulation test results indicate that the null hypothesis of no manipulation cannot be rejected[3].

[3] The manipulation tests are robust to the bandwidth selected or recommended via the data-driven process.

Figure 1

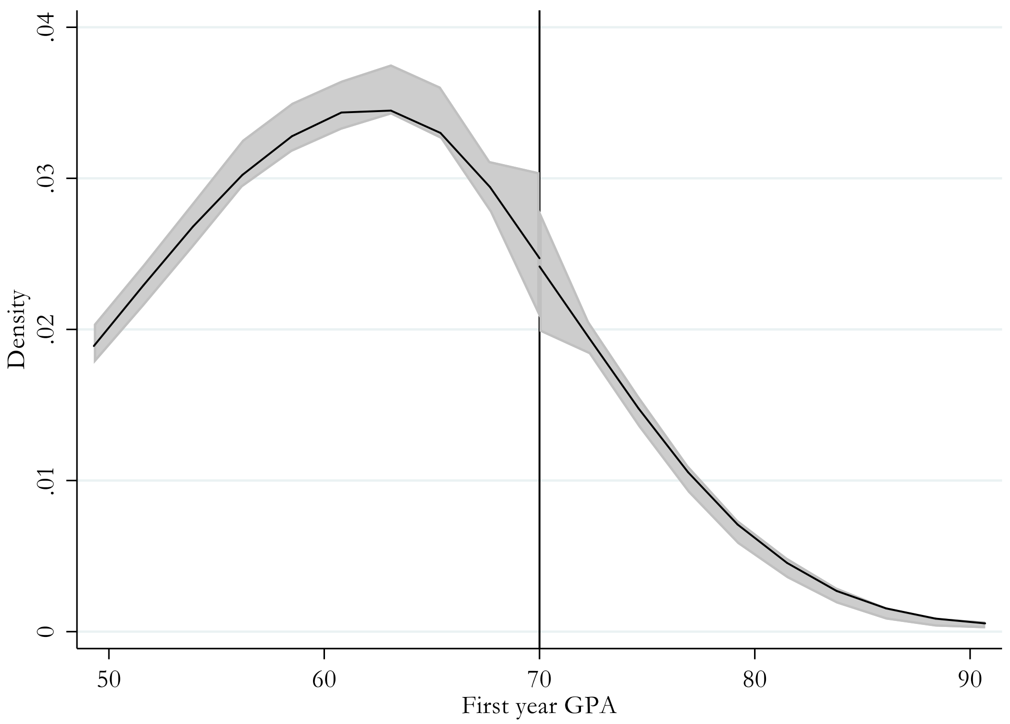

Figure 2: Dean's Merit List Density Test, whole sample

The density tests by faculty presented in Figure 2 largely support the causal interpretation of the RD results. Figures 1 and 2 contain no statistically significant discontinuities at any of the bandwidths examined. The student GPA is given as a number between 0 & 100, with the average student achieving a GPA in the mid 50's to mid-60's.

Figure 2

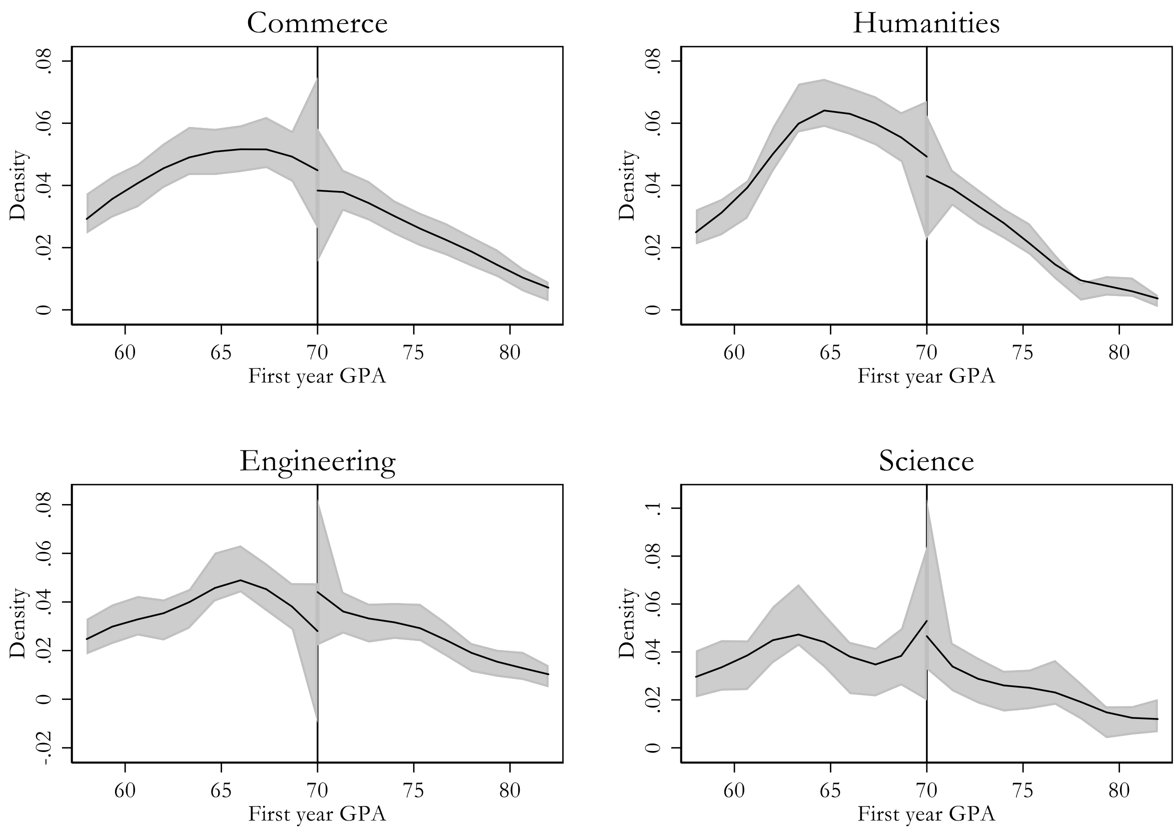

Figure 2: Density tests by Faculty

From our perspective, based on the grading system, students who achieve an average of 70 to 75% are inherently different to students who achieve GPA's above 75%. Importantly, 70-74% represents an upper-second class of pass while 75% and above represents a first-class pass. The RD manipulation tests also indicate that it would be sensible to conduct the analysis with varying bandwidth sizes to ensure robustness of results. Therefore all results presented will commence with the CJM optimal bandwidth, followed by researcher-selected bandwidths of BW(2) = 4.8, BW(3) = 4, BW(4) = 3 and BW(5) = 2[4].

[4] CJM is representative of the optimal bandwidth given by the Cattaneo, Jannson and Ma (2018) data-driven methodology.

A potential complicating issue that arises from the choice of bandwidth is explanatory power of the RD model. The model relies on a large number of observations around the cutoff to ensure validity of the results. Using individual cohorts will lead to small numbers, however the cohorts are combined in this analysis, giving us large sample sizes had we instead selected only one entering cohort. By ensuring we have our researcher-selected bandwidths in addition to the CJM-selected bandwidths, we 1) ensure we have sufficient observations to account for any statistical noise that might occur and 2) we do not introduce confounding effects into the data by venturing into different classes of pass.

Further tests were conducted to ensure other characteristics are continuous through the threshold. To do this, the procedure from Casey et. al (2018) was followed, by regressing predetermined student characteristics on GPA. If substantial discontinuities exist at the DML threshold, these might invalidate the validity of the research design, indicating that individuals with certain characteristics are non-randomly allocated around the threshold.

The results for the manipulation tests are presented in Table 1. Using a set of predetermined characteristics, we find there are no statistically significant differences across the DML threshold. Importantly, a strong predictor of performance such as matric performance (entrance score) is not discontinuous across the DML threshold.

Another potential complicating factor is the re-enrolment of students in subsequent years. This could affect the RD estimates if students above the threshold dropped out at significantly higher rates than those with GPA's below the threshold. However, evidence suggests this is not the case and so does not present a significant threat to the validity of the results.

Table 1

Validity of RD Design - Balance on Covariates

	(1)	(2)	(3)	(4)	(5)
Covariates	**BW=CJM**	**BW = ± 4.8**	**BW = ± 4**	**BW = ± 3**	**BW = ± 2**
Female	1.2395	3.1436*	2.1884	1.7085	0.2944
	[0.8724]	[1.1618]	[1.4714]	[2.2429]	[3.9321]
African	0.8432	1.4720*	0.3765	-0.7055	-1.0617
	[0.6651]	[0.8888]	[1.1450]	1.7554	[3.0769]
Coloured	-0.6899	-0.4033	0.0937	0.5944	2.6837
	[0.5767]	[0.7654]	[0.9529]	[1.4451]	[2.5719]
Indian/Asian	0.2094	0.2976	0.8915	2.6527	1.0064
	[0.4267]	[0.5699]	[0.7133]	[1.0821]	[2.0664]
Humanities	0.5144	1.4253	1.4157	1.0441	2.0929
	[0.8331]	[1.1105]	[1.4095]	[2.1584]	[3.7905]
Science	-0.0613	0.43773	0.0014	1.0886	4.1048
	[0.5614]	[0.7411]	[0.9509]	[1.4298]	[2.5511]
Engineering	-1.2748*	-1.2347	-0.0643	-0.5248	-0.4901
	[0.6607]	[0.8739]	[1.0863]	[1.6608]	[2.9108]
3-year degree	0.2154	0.2227	-0.0460	0.5189	3.8801
	[0.8649]	[1.1529]	[1.4604]	[2.2301]	[3.9233]
Private school	-0.2187	0.0661	0.76914	-0.5522	2.5633
	[0.8472]	[1.1277]	[1.4279]	[2.1896]	[3.8311]
House of Delegates	-0.2736	-0.2621	-0.2405	0.9573	1.7376
	[0.2934]	[0.3804]	[0.4755]	[0.6772]	[1.2365]
House of Representatives	0.1768	0.4874	0.3289	-0.3173	-0.0941
	[0.4104]	[0.5481]	[0.7088]	[1.1065]	[1.9182]
Department of Education and Training	0.1803	0.7376]	0.2456	0.1603	-0.2895
	[0.4538]	[0.5964]	[0.7764]	[1.1875]	[2.1223]
Entance Score	7.6278	6.0291	20.8213	33.8186	139.2332*
	[14.6206]	[19.8623]	[25.3932]	[38.3141]	[69.5935]
Financial Aid status	-0.00415	0.3966	-0.0780	-0.4964	-2.298
	[0.4749]	[0.6344]	[0.8112]	[1.2379]	[2.2085]
Residence Status	-1.0923	-1.7822*	-2.0648	-2.2567	1.6271
	[0.8093]	[1.0745]	[1.3609]	[2.0695]	[3.5892]
Extended Programme	0.7341	0.4132	0.1251	0.4137	-0.4502
	[0.4963]	[0.6721]	[0.8609]	[1.3167]	[2.3688]
Observations	2153	1989	1686	1264	858

Notes: This table presents estimates of the above-cutoff indicator using predetermined covariates as the dependent variable. Column (1) uses observations within the optimal CJM bandwidth. Column (2) uses our preferred bandwidth based on the grade structure at the institution under consideration. Columns (3), (4) and (5) each use a fixed bandwidth of 4, 3 and 2 grade points around the cutoff, respectively. All estimates are clustered on the running variable.

*p value <0.1

5 Results

5.1 Short term results

The evidence presented in the previous section is indicative of the validity of the RD design in this context. In addition, the tests run of the predetermined characteristics on the treatment variables showing that the predetermined characteristics are not discontinuous through the DML threshold are further support for the RD design used in this paper. This allows for the interpretation of any associated discontinuities in student outcomes as the causal effect of academic recognition at the end of the first year of study.

We now estimate student outcomes for students marginally above the threshold versus those marginally below the threshold[5]. Table 2 shows the short-term outcomes of academic recognition on students' annual non-cumulative GPA for the sample as a whole. On average the effect of the DML policy is negative with the impact quite significantly more negative in the third year of study compared to the second. Depending on the classification of third year GPA relative to first year, this is evidence that the policy has lasting effects on student performance. Overall, students treated by the DML policy in the first year appear to experience a less stable academic pathway as the impact of the policy is negative and lasting. Importantly, robustness checks are included in each set of results presented in this section, allowing us to observe the robustness checks at the same time as viewing the main results. Irrespective of the chosen bandwidth, the current implementation of the DML policy leads to lasting, negative impacts on treated students relative to those not treated by the policy.

[5] We continue to estimate the impact of academic recognition on all five identified bandwidths relative to the DML threshold.

Table 2

	(1)	(2)	(3)	(4)	(5)
Annual non-cumulative GPA Year2	**BW=CJM**	**BW = ± 4.8**	**BW = ± 4**	**BW = ± 3**	**BW = ± 2**
Robust	-1.058*	-1.388**	-1.668***	-2.387***	-2.884***
	[0.544]	[0.575]	[0.615]	[0.669]	[0.718]
Conventional	-1.047***	-0.781*	-0.908**	-1.154**	-1.851***
	[0.384]	[0.415]	[0.447]	[0.505]	[0.586]
Observations	2153	1864	1539	1148	753
Annual non-cumulative GPA Year3					
Robust	-2.690***	-3.153***	-3.517***	-4.484***	-4.868***
	[0.911]	[0.988]	[1.06]	[1.172]	[1.298]
Conventional	-1.735***	-2.041***	-2.274***	-2.697***	-3.786***
	[0.632]	[0.707]	[0.769]	[0.875]	[1.034]
Observations	2144	1857	1537	1148	751

Estimates are presented for the full sample. Standard errors are clustered along the running variable. * implies p value<0.1, ** implies P < 0.05, and *** implies P < 0.01. This specification does not include covariates.

Table 2 shows that the closer students are to the cutoff, the greater the disincentive effect in the subsequent years. Even though the CJM result must be interpreted with caution as it crosses over into a new class of pass above the threshold, the results are nonetheless negative and significant, and fairly close to the next cutoff at 4.8% above and below the threshold. The next step is to split the sample into groups by degree duration. This splits the sample into two cohorts, those students registered for 3-year degrees and those registered for 4-year options.

Figure 3

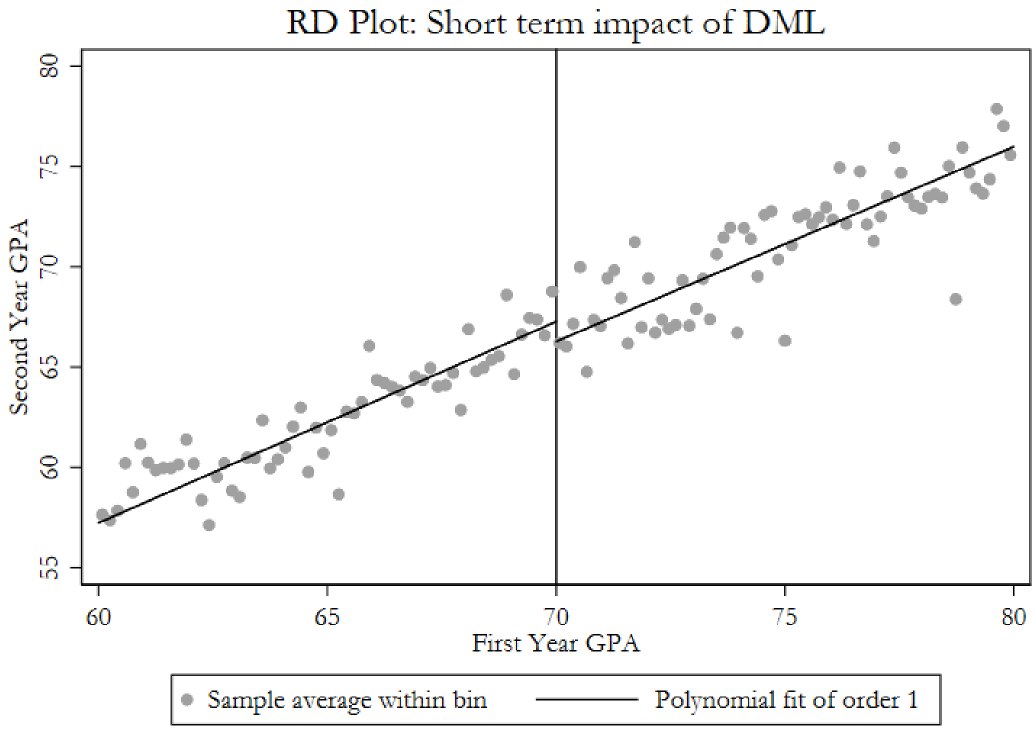

Figure 3: RD plot of treatment effect

Figure 3 is a visual presentation of the top set of results from Table 2, specifically presenting the results from column (1). The vertical distance at the cutoff is exactly equal to the value of the robust estimator of -1.058, indicating that students who are recognized with the DML in year one perform on average worse in year 2 relative to those students who were statistically similar but were not recognized with a DML award.

Table 3

	(1)	(2)	(3)	(4)	(5)
Annual non-cumulative GPA Year2	**BW=CJM**	**BW = ± 4.8**	**BW = ± 4**	**BW = ± 3**	**BW = ± 2**
Robust	-2.066**	-2.389**	-2.797**	-4.285***	-5.0118***
	[0.958]	[1.061]	[1.150]	[1.280]	[1.419]
Conventional	-1.405**	-1.358*	-1.5667*	-1.914**	-3.096***
	[0.642]	[0.735]	[0.802]	[0.923]	[1.127]
Observations	1229	1066	881	652	422
Annual non-cumulative GPA Year3					
Robust	-3.472**	-3.926**	-4.279**	-5.331***	-2.543***
	[1.385]	[1.542]	[1.697]	[1.920]	[2.085]
Conventional	-2.229**	-2.701***	-2.97***	-3.430***	-4.475***
	[0.883]	[1.004]	[1.115]	[1.329]	[1.679]
Observations	1220	1059	877	648	418

Table 3: Impact of the Dean's Merit List Policy: 3-year degree. Estimates are presented for the full sample. Standard errors are clustered along the running variable. * implies p value<0.1, ** implies P < 0.05, and *** implies P < 0.01. This specification does not include covariates.

For the 3-year degree options, the DML has a significantly negative effect both when measured using the CJM bandwidth and the other, smaller bandwidth options. Students treated with the DML experience a significantly lower annual GPA in subsequent years compared to their first year of enrolment. In particular, the effect is about 50% larger in the third year of study relative to the second. This holds across all potential bandwidths and is especially pronounced in columns (4) and (5). The treated students do significantly worse compared to similar students who were not treated.

Table 4

	(1)	(2)	(3)	(4)	(5)
Annual non-cumulative GPA Year2	**BW=CJM**	**BW = ± 4.8**	**BW = ± 4**	**BW = ± 3**	**BW = ± 2**
Robust	0.486	0.213	0.110	0.597	1.107
	[1.061]	[1.130]	[1.209]	[1.312]	[1.466]
Conventional	-0.494	0.109	0.131	0.053	0.056
	[0.749]	[0.812]	[0.875]	[0.994]	[1.146]
Observations	924	798	658	496	331
Annual non-cumulative GPA Year3					
Robust	-1.411	-1.798	-2.084	-2.523	-2.316
	[1.334]	[1.401]	[1.460]	[1.580]	[1.856]
Conventional	-1.002	-0.9777	-1.179	-1.500	-2.362*
	[0.999]	[1.091]	[1.159]	[1.267]	[1.397]
Observations	924	798	658	496	331
Annual non-cumulative GPA Year4					
Robust	-1.555	-1.251	-1.482	-1.249	-3.159**
	[1.26]	[1.331]	[1.371]	[1.408]	[1.421]
Conventional	-2.299	-2.020**	-1.731	-1.713	-1.505
	[0.926]	[1.012]	[1.089]	[1.202]	[1.297]
Observations	800	687	567	433	287

Table 4: Impact of the Dean's Merit List Policy: 4-year degree. Estimates are presented for the full sample. Standard errors are clustered along the running variable. * implies p value<0.1, ** implies P < 0.05, and *** implies P < 0.01. This specification does not include covariates.

The results for the 4-year degree students differ substantially to the 3-year degree students. Overall the DML does not have an impact on subsequent student GPA. The differences in entry characteristics of the 4-year students relative to the 3-years students may account for the observed differences in the impact of the DML on student performance. Students who enroll for the 4-year degrees tend to have higher matric exam scores (entry scores), complete more subjects in matric and are more likely to sign-up for math-based disciplines such as actuarial science and engineering. There may also be unobservable differences between the two cohorts across both degree types that could explain the very different impacts on the non-cumulative annual GPA.

Further analysis was conducted to evaluate DML policy effects by gender and race. There are surprisingly little effects by race which remains difficult to explain. The results for an analysis based on gender reveals significant differences in response by males and females. For most of the female sub-samples, results are not statistically significant. The opposite is true for males. Results for males are large and significant, indicating gender as one of the key differences driving the heterogeneity observed within the findings.

Based on the rigid degree requirements across the university it is unlikely that students start engaging in strategic course selection behavior from year 2 onward to maintain their DML status. While first year curricula are fairly fixed, second year curricula are marginally more relaxed by allowing students to elect at most 2 of their courses out of a maximum of 9. This does not give students enough room to manipulate their GPA as the elective courses are typically lower weighted relative to compulsory courses for either the degree or specialisation in a discipline.

The significant negative results observed in this analysis are supported by evidence from the psychology literature. Eisenberger and Cameron (1996) show that individuals given quality-dependent rewards, where quality may be measured by achieving some minimum benchmark or score, could show increased perceptions of competence by recipients but not necessarily reductions in intrinsic motivation. In line with this, Sternberg and Lubart (1996) show that extrinsic motivators such as reward systems are desirable and may have positive impacts if implemented at certain stages of effort but not all stages of work.

Overall, the results are suggestive that the DML recognition does not have the desired impact at the institution under study. The overall effect of the policy is negative, and within the degree types the two groups of students appear to be reacting differently to the implementation of the policy.

6 Conclusion

While many colleges and universities have policies that recognize good student performance, very little is known about the impact these policies have on student outcomes. To the best of our knowledge, we are aware of only a few papers in the economics literature that examine this policy, with no literature from Europe, Asia or South America. Therefore it is difficult to locate the results of this study in the context of the broader literature of higher education systems around the world.

This paper evaluated the impact of academic recognition policies on student outcomes. Using a regression discontinuity approach, evidence is provided to show that a DML policy at first year level does not yield the desired response from students. Moreover, the response from students intensifies over time, leading to significantly lower performance in subsequent years from students who are expected to be top performers in their respective programmes.

While the results presented in this paper are largely discouraging, it does represent an opportunity for university administrators to rethink the student incentive structure and to come up with the system that takes student motivators and beliefs into account.

References

Bénabou, R. and Tirole, J., 2003. Intrinsic and extrinsic motivation. *The review of economic studies*, *70*(3), pp.489-520.

Bettinger, E.P. and Baker, R.B., 2014. The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, *36*(1), pp.3-19.

Casey, M.D., Cline, J., Ost, B. and Qureshi, J.A., 2018. Academic Probation, Student Performance, and Strategic Course-Taking. *Economic Inquiry*, *56*(3), pp.1646-1677.

Cattaneo, M.D., Jansson, M. and Ma, X., 2018. Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), pp.234-261.

Cook, T.D and Campbell, D.T., 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin.

Crandall, V. and McGhee, P., 1968. Expectancy of reinforcement and academic competence. *Journal of Personality, 36,* pp.635-648.

Eisenberger, R. and Cameron, J., 1996. Detrimental effects of reward: Reality or myth?. *American psychologist*, *51*(11), p.1153.

Fletcher, J.M. and Tokmouline, M., 2010. The effects of academic probation on college success: Lending students a hand or kicking them while they are down?

Goldrick-Rab, S.N., Harris, D., Kelchen, R. and Benson, J., 2012. Need-based financial aid and college persistence experimental evidence from Wisconsin.

Imbens, G.W. and Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, *142*(2), pp.615-635.

Lee, D.S. and Card, D., 2008. Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), pp.655-674.

Lee, D.S. and Lemieux, T., 2009. Regression discontinuity designs in economics (No. w14723).

Leeds, D.M. and DesJardins, S.L., 2015. The effect of merit aid on enrollment: A regression discontinuity analysis of Iowa's National Scholars Award. *Research in Higher Education*, *56*(5), pp.471-495.

Lindo, J.M., Sanders, N.J. and Oreopoulos, P., 2010. Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, *2*(2), pp.95-117.

Massey, D.S. and Mooney, M., 2007. The effects of America's three affirmative action programs on academic performance. *Social Problems*, *54*(1), pp.99-117.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), pp.698-714.

Seaver, W.B. and Quarton, R.J., 1973. Social Reinforcement of Excellence: Dean's List and Academic Achievement.

Sternberg, R.J. and Lubart, T.I., 1996. Investing in creativity. *American psychologist*, *51*(7), p.677.

Wright, N., 2018. Perform Better, or Else: Academic Probation, Public Praise and Students Decision-Making. *Andrew Young School of Policy Studies Research Paper Series*.