Thesis

Sophia Yang

Introduction

In the wake of the COVID-19 pandemic, the Trinity College of Arts and Sciences at Duke University introduced academic policy changes. During the Spring 2020 semester, all courses were switched to Satisfactory/Unsatisfactory (S/U) grading as opposed to the traditional graded (A- F) scale. While this was the only semester where all courses were allowed to be S/U and count for all requirements, other policy changes during the pandemic regarding S/U still remain.

Referencing Duke University Undergraduate Bulletins, prior to Spring 2020, S/U grading was only permitted to count towards total graduation credits. After Spring 2020, S/U grading has additionally been permitted to count towards general education requirements. Some faculty members have noted an increase in student usage of S/U and departments are moving towards assigning courses to be graded (A-F) only. The most recent of many changes has been a switch in defaults so instructors have to opt-in for allowing S/U rather than having to specifically opt-out of allowing S/U[1]. With the rollback of most pandemic-era policies such as masking and social distancing long gone, this raises the question of whether the S/U policy should return to its pre-pandemic form.

A growing body of literature has examined the impact of the pandemic on student learning and a plethora of studies have been done on pass/fail grading in post-secondary education (although primarily in medical school). Rather than exploring impacts on student outcomes, this thesis seeks to determine how the changes in flexible grading policy have impacted student behaviors with the goal of better informing educators and administrators.

[1] https://dukechronicle.com/article/duke-university-arts-and-sciences-council-changes-satisfactory-unsatisfactory-grading-system-duke-athletics-changing-landscape-20250503

Background and Context

Birth of Traditional A-F Grading

Education is notably missing from the US Constitution, but in the period between 1852 to 1918 all states had passed legislation requiring compulsory education (Diorio 2023)[2]. As a result, K-12 enrollments nearly tripled from 1870 to 1910 (Goldin 2006)[3]. Simultaneously, the Morrill Acts of 1862 and 1890 provided federal land for the establishment of public colleges, opening the doors for access to higher education. Historically, students were evaluated by descriptions and evaluations given by individual teachers. Even entrance exams for college were made by individuals and occurred at individual schools, making results unreliable (Schudson 1972; Wechsler 1977)[4][5]. The massive increase in the number of students required a revolutionary new approach to education that could scale with the increasing demand for education. A standard grading system was needed.

The movement towards the usage of report cards and grades as a success indicator became widespread. Instead, the rise of national examinations such as those by the College Board arose (Valentine 1987)[6]. Now, the shift was towards more consistent report cards across a school or district as a success indicator. Additionally, universities began using academic "credits" to quantify the amount of work and subsequently the amount of effort a student took on during a given period. Combined with grades, one could compare a student not just within their class but also with students in other classes in the same school. However, grading systems remained variable across the country in regards to what to measure and their frequency (Ashbaugh and Chapman 1925)[7]. Many debates arose ranging from the potential misinterpretation of grades (Bixler 1936)[8] to the discrepancy between standardized tests and teacher's grades (Hadley 1954)[9] to the balance of extrinsic and intrinsic motivation (Sumner 1935)[10]. Yet despite

[2] https://www.ebsco.com/research-starters/history/history-public-education-us

[3] Goldin, Claudia , "School enrollment and pupil–teacher ratios, by grades K–8 and 9–12 and by public–private control: 1869–1996." Table Bc7-18 in Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. New York: Cambridge University Press, 2006. http://dx.doi.org/10.1017/ISBN-9780511132971.Bc1-509

[4] Schudson, M. S. (1972) Organizing the meritocracy. Harvard Educational Review, 42(1), 34-69. https://eric.ed.gov/?id=EJ056906

[5] Wechsler, H. S. (1977) The Qualified Student: A History of Selective College Admission in America, 1870-1970 (New York, NY: Wiley-Interscience). https://eric.ed.gov/?id=ED150933

[6] The College Board and the School Curriculum. A History of the College Board's Influence on the Substance and Standards of American Education, 1900-1980. https://eric.ed.gov/?id=ED285443

[7] Ashbaugh, E. J. and Chapman, H. B. (1925) Report cards in American cities. Educational Research Bulletin, 4(14), 289-293.

[8] Bixler, H. H. (1936) School marks. Review of Educational Research, 6(2), 169-173.

[9] Hadley, S. Trevor (1954) A school mark-fact or fancy. Educational Administration and upervision, 40, 305-312.

[10] Sumner, R. G. (1935) What price marks?. Funior-Senior High School Clearing House, 9(6), 340-344.

the multitude of flaws grading had, the rapidly growing demand of education necessitated a solution.

During the early 1900s, research was conducted to determine the best way to assign grades. Studies such as that by (Starch 1913)[11] found that using a 100 percent scale was highly inconsistent across teachers. In response, some suggested a categorical system of "diagnostic letters" to reduce the impact of inconsistency on reported grades (Finkelstein 1913)[12]. By the 1940s, the A-F grading system was adopted by over 80% of U.S. schools, rising in popularity along with the 4.0 scale (Schneider and Hutt 2013)[13].

Rise of Pass/Fail

While A-F grading rose to prominence, its core problems remained. Grades were still inaccurate, being assigned differently across instructors, departments, and institutions. Additional studies found poor correlations between college grades and post-educational success (Hoyt 1966)[14]. Concerns about student learning under A-F grading were also raised. As put by Stallings and Leslie,

> The undergraduate perceives grades as that proverbial sword hanging over his head which forces him to study content he otherwise might not study. The power of 'the grade' is strong enough to restrict his studying to material which he anticipates will be on tests… (Stallings and Leslie 1970)[15].

Criticism of traditional A-F grading led to an era of educational innovation. Many schools began experimenting with alternate forms of grading, the most prominent of which was the pass/fail system. Pass fail grading was not a new idea, with records in American higher education from as early as 1851 (Smallwood 1935)[16]. However, it remained obscure until the 1960s and 1970s. Proponents of pass fail grading argued that it would foster an intrinsic interest in learning and greater exploration of academic courses. As put by Weller, it was hoped to "free the instructor and the student to communicate on a colleague to colleague

https://www.jstor.org/stable/30174436

[11] Starch D (1913). Reliability and distribution of grades. Science 38, 630-636. Medline, Google Scholar https://doi.org/10.1126/science.38.983.630

[12] Finkelstein IE (1913). The Marking System in Theory and Practice, Baltimore: Warwick & York. https://books.google.com/books?id=4wsRAAAAMAAJ&lr=&source=gbs_navlinks_s

[13] Schneider, J., & Hutt, E. (2013). Making the grade: a history of the A–F marking scheme. *Journal of Curriculum Studies*, *46*(2), 201–224. https://doi.org/10.1080/00220272.2013.790480

[14] Hoyt, D. P. (1966). College grades and adult accomplishment: A review of research. *The Educational Record*, 47, 70-75.

[15] Stallings, W. M., & Leslie, E. K. (1970). Student attitudes towards grades and grading. Improving College and University Teaching, 18(1), 66-68 https://files.eric.ed.gov/fulltext/ED060054.pdf

[16] https://books.google.com/books/about/An_Historical_Study_of_Examinations_and.html?id=OMgjAAAAMAAJ

basis" (1983)[17]. By the early 1970s, no penalty grading was present in some capacity in over two-thirds of a sampled 2500 American colleges and universities (Elsner and Brydon 1974)[18].

Pass fail grading was not without faults. Multiple studies of the time found that pass/fail grading was often used to concentrate more effort in graded classes to boost or maintain grade point averages (Quann 1971; Collins and Nickel 1975)[19][20]. Whether this encouraged exploration outside the major is unclear with both positive (Sgan 1969)[21] and negative (Johnson 1970; Weems et al. 1971)[22][23] reports. What was apparent was that students using pass/fail were less engaged in course material than their graded counterparts. In a study by Karlins et al, traditionally graded students reported completion of 80% of readings and 85% attendance as opposed to pass/fail students completion of 61% of readings and 74% attendance (Karlins et al. 1969)[24]. Critics also argued against the binary extremes of pass/fail as well as highlighting administrative challenges regarding dean's list, calculation of grade point averages, and transfer students. Schools thought that they needed traditional grades to motivate students and that grades convey important information about a student to future employers or higher level educational institutions. Regarding the intent to create bonds between instructors and students, Weller (1983)[25] found that pass/fail grading did not increase faculty evaluation time and institutions were divided on if it had a positive impact on faculty evaluation of students. Nearly 2 to 1 of the pass/fail institutions surveyed believed pass/fail grading did not result in a more positive student perception of grading.

Decline of Pass/Fail

While research had been conducted on education, the issue of education had largely remained out of the public eye until the 1980s. It was common belief that schools did not matter, and this was given scientific backing by the 1966 Coleman report which found that family background was more influential to student achievement than schools themselves (Coleman et al. 1966)[26]. As a result, a relaxed attitude towards academics was commonplace and had provided the backdrop for introducing pass/fail.

[17] https://eric.ed.gov/?id=EJ288937
[18] https://eric.ed.gov/?id=ED088549
[19] https://files.eric.ed.gov/fulltext/ED051737.pdf
[20] Collins, J. R., & Nickel, K. N. (1975). Grading policies in higher education: the Kansas Study/ the National survey. University Studies, 103, Wichita.
[21] https://doi.org/10.2307/1979547
[22] Johnson, Jack T. Evaluate program, not grading. College and University Business, 49(3), 1970, 77-78.
[23] Weems, John E., Clements, William H., Quann, Charles J., Smith, K., and Schefelbein, Barbara E. Pass -fail; Were the hypotheses valid? College and University, 46, 1971, 535 -556.
[24] Karlins, Marvin, Kaplan, Martin, and Stuart, William. Academic attitudes and performance as a function of differential grading systems: An evaluation of Princeton's pass-fail system. Journal of Experimental Education, 37(3), 1969, 33-50.
Available online at https://www.tandfonline.com/doi/permissions/10.1080/00220973.1969.11011129?scroll=top
[25] https://eric.ed.gov/?id=EJ288937
[26] https://eric.ed.gov/?id=ED012275

However, newly emerging research was beginning to suggest that schools did matter. In response to Coleman, the effective schools movement sought to analyze characteristics of schools that correlate with higher academic achievement. Edmonds (1979) expanded upon prior studies such as that of Weber (1971)[27] to analyze practices used by schools with high performing students and outlined characteristics of effective schools[28]. As Edmonds put it, "We can, whenever and wherever we choose, successfully teach all children whose schooling is of interest to us. We already know more than we need to do that. Whether or not we do it must finally depend on how we feel about the fact that we haven't so far"[29]. His work was later expanded upon with additional and refined correlates of effective schools by others including Lezotte (1991)[30]. Independent research from the UK by Rutter et al. (1982)[31] further strengthened the case for better schools.

In 1983, the National Commission on Excellence in Education published A Nation at Risk. In this monumental report, researchers found consistent declines in high school and college student achievement scores and recommended high school graduation requirements (Gardner et al. 1983)[32]. The report describes the state of American education as "unilateral educational disarmament" and warns of a "rising tide of mediocrity", capturing media attention across the country. Overnight, education became a nonpartisan issue. Pamphlets from the Department of Education made research more accessible to the public (1986)[33], the National Board for Professional Teaching Standards was established, exams began to shift away from multiple choice questions, and the first education summit of the nation's governors was held (Ravitch 1990)[34]. As a result of the growing importance of education to the general public, schools largely returned to a system of traditional A-F grading. The binary nature of pass/fail grading obscured the student data necessary to measure student achievement and improvement of the education system.

Rising Educational Attainment

Educational attainment in America rose sharply in the mid to late 1980s as college degrees became increasingly important. Papers from economists found that education was a way

[27] https://eric.ed.gov/?id=ED057125

[28] https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_197910_edmonds.pdf

[29] https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_197910_edmonds.pdf

[30] Lezotte, L. W., Correlates of Effective Schools: The First and Second Generation https://www.effectiveschools.com/Correlates.pdf

[31] Fifteen Thousand Hours; Rutter, et al, Harvard University Press, Cambridge, MA ISBN 9780674300262 https://www.hup.harvard.edu/books/9780674300262

[32] https://eric.ed.gov/?id=ED226006

[33] What Works. Research about Teaching and Learning. Department of Education, Washington, DC.; Office of Educational Research and Improvement (ED) https://files.eric.ed.gov/fulltext/ED263299.pdf

[34] Education in the 1980's: A Concern for 'Quality' Diane Ravitch — January 10, 1990 https://www.edweek.org/policy-politics/opinion-education-in-the-1980s-a-concern-for-quality/1990/01

to signal and screen for high-ability workers (Stiglitz 1975)[35]. Enticed by the promise of employment, more Americans obtained post-secondary degrees. According to data from the US Department of Education National Center for Education Statistics, the percentage of American adults aged 25 or older who held at least a Bachelor's degree continuously rose from 6.2% in 1950 to 25.6% in 2000 to 37.5% in 2020 (2025)[36]. A significant source of this increase was the rise of for-profit "diploma mills". Fueled by the exploitation of financial aid and the political climate of deregulation and privatization, education became an industry. In 1990, there was only a single publicly traded for-profit university but by 2000 there were 40 publicly traded for-profit universities (Beaver 2017)[37]. The Senate Committee on Health Education, Labor and Pensions found that "Between 1998 and 2008, enrollment at for-profit colleges increased 225 percent, compared to 31 percent growth in higher education" (2012)[38].

The goal of these for-profit institutions is to cut costs and grow profits. To do so, these institutions prioritized enrollment over teaching. They employed 10 times the recruiters for every career-service employee and hired mostly part-time staff (Senate Committee on Health, Education, Labor, and Pensions 2011)[39]. As a result, the outcomes of students at these institutions has been subpar. Data from the Department of Education suggests that most for-profit career programs fail to benefit students with 72% of programs having graduates earning less than high school dropouts, compared to 32% at public institutions (2012)[40].

The Price of Education

In the emerging "credential society" social classes were distinguished by the degree which one held and the prestige associated with that school (Collins 1979)[41]. Degrees from elite schools acted as insurance for the future against rising the "fear of falling" of the middle class as household wealth inequalities rose (Ehrenreich 1989)[42]. Backed by impressive scholars and research contributions, admission into these elite colleges has always been challenging. Now, with the over saturation of degree holders, many would pay whatever price necessary for prestige to ensure financial stability.

At first, institutional rankings originated from athletic college affiliation (Ivy League) and regional primacy (e.g. Duke in the South, USC and Stanford in the West). By the end of the

[35] Stiglitz, J. E. (1975). The Theory of "Screening," Education, and the Distribution of Income. The American Economic Review, 65(3), 283–300. http://www.jstor.org/stable/1804834

[36] Hanson, Melanie. "Education Attainment Statistics" EducationData.org, 2025-01-14, https://educationdata.org/education-attainment-statistics

[37] Academe, Vol. 103, Number 1, in the January-February 2017 https://www.aaup.org/academe/issues/103-0/rise-and-fall-profit-higher-education

[38] https://www.help.senate.gov/imo/media/for_profit_report/PartI.pdf

[39] https://www.help.senate.gov/imo/media/for_profit_report/ExecutiveSummary.pdf

[40] https://www.ed.gov/sites/ed/files/policy/highered/reg/hearulemaking/2012/notice-proposed-rulemaking-march-14-2014.pdf

[41] https://archive.org/details/credentialsociet0000coll/page/n5/mode/2up

[42] https://archive.org/details/fearoffalling00barb

20th century, third-party ranking systems like that from U.S. News beginning in 1983[43]. A dominant emergent strategy in the battle for the best students has been raising tuition and offering lucrative scholarships to high-achievers.

States have cut funding for public universities since the 1980s. This only accelerated in the 21st century. In response to the Great Recession (2007-2009) and mandatory spending programs like Medicaid, between 2008 and 2013 appropriation for the median public university declined by over 20% per full-time student[44]. As a result, many public institutions were forced to increase tuition. A decade after the recession, state funding for higher education has not rebounded in most states[45].

According to analyses by Banks et al. (2024)[46], annual tuition and fees at private 4-year institutions during the 1979-1980 academic year was $11,357 (adjusted for inflation), compared to the $2,599 (adjusted for inflation) at public institutions. Over time this gap has widened to a difference of over $20,000 by 2019-2020. By the 2019-2020 academic year, average annual tuition and fees at both public and private 4-year institutions had risen nearly 3 times the cost in 1979-1980, adjusted for inflation. Without adjusting for inflation, the cost of higher education has jumped 10-fold.

Grade Inflation

As the institutions changed, so did the students. Seeking to distinguish themselves from the increasing number of degree holders and limited by rising tuition, students sought to maximize their grade point averages (GPA) for future profit rather than of pure educational interest. The "entrepreneurial student" shopped "for bargain courses, encouraged by a faculty whose jobs are defined by "course load", administrators who deal in credit hours as if they were coin, [and] institutions whose corpus evolves steadily into the corporate" (Haswell 1999)[47]. Yet, not all gains in GPA necessarily match skill.

During the Vietnam War (1955–1975), college enrollment was used to avoid the draft. As a result, failing a student could directly result in their conscription. Evidence has shown an increase in grading leniency due to this policy (Bejar and Blew 1981; Birnbaum 1977)[48][49]. In

[43] https://www.usnews.com/rankings

[44] https://www.amacad.org/sites/default/files/academy/multimedia/pdfs/publications/researchpapersmonographs/PublicResearch

[45] https://www.cbpp.org/research/state-budget-and-tax/unkept-promises-state-cuts-to-higher-education-threaten-access-and

[46] https://law.stanford.edu/stanford-center-for-racial-justice/projects/private-universities-in-the-public-interest/private-universities-in-the-public-interest-white-paper/

[47] https://www.jstor.org/stable/359043

[48] Bejar, I. L. and Blew, E. O. (1981) Grade inflation and validity of scholastic aptitude test. *American Educational Research Journal*, 18, 143–156. https://doi.org/10.3102/00028312018002143

[49] Birnbaum, R. (1977) Factors related to university grade inflation. Journal of Higher Education, 5, 519–538. https://doi.org/10.1080/00221546.1977.11776572

a study by Rojstaczer and Healy (2012)[50] they found, "in 1960, as in the 1940s and 1950s, C was the most common grade nationwide; D's and F's accounted for more grades combined than did A". By the end of the Vietnam War, As and Bs made up half to two thirds of grades in American colleges (Davidson 1975)[51]. After the conclusion of the Vietnam War, grades remained a measure of more than a student's academics. Grades were affected by all manner of things from a teacher's concern about student self-esteem, departmental policy to attract students, and the impact of grades during job search (Schneider and Hutt 2013)[52].

One of the most predominant reasons for grade inflation was the rise of student evaluation of teaching (SET). SET first began to rise in popularity alongside the Civil Rights movement as a way for students to voice their complaints (Valsan and Sproule 2008)[53]. Under the belief that student evaluations measure teaching effectiveness, administrators realized the opportunity evaluations presented to advertise their programs with some universities going as far as using evaluations as a component of consideration for promotion, tenure, and resource allocation. By 1980s, SET became commonplace in American higher education (Centra 1993; Wachtel 1998)[54][55]. Yet, "the typical SET questionnaire treats the student as a customer and measures the satisfaction of the student with his or her professor, and not learning" (Crumbley 2010)[56]. Multiple studies have found significant, positive correlations between student evaluations and student grades (Langbein 2008; Ellis 2003)[57][58]. On the other hand, SET rankings are not significantly correlated with actual student learning (Uttl 2017)[59]. An article by Neath titled "How to Improve Your Teaching Evaluations Without Improving Your Teaching" even goes so far as to suggest multiple methods such as getting evaluated before exams and grading leniently (Neath 1996)[60]. Simultaneously, efforts to cut costs and increase profit margins resulted in a rise of nontenured, adjunct faculty (Bettinger and Long 2010)[61]. These educators' careers depended significantly on SET rankings. As such, over time, professors became increasingly aware of the implications SET rankings could have on their careers.

Grade inflation is unevenly applied across institutions and subjects. Average student GPAs in

[50] https://journals.sagepub.com/doi/abs/10.1177/016146811211400707

[51] Davidson, J. F. (1975) Academic interest rates and grade inflation. Educational Record, 56, 122–125. https://eric.ed.gov/?id=EJ122844

[52] Schneider, J., & Hutt, E. (2013). Making the grade: a history of the A–F marking scheme. *Journal of Curriculum Studies*, *46*(2), 201–224. https://doi.org/10.1080/00220272.2013.790480

[53] Valsan, C., & Sproule, R. (2008). The Invisible Hands behind the Student Evaluation of Teaching: The Rise of the New Managerial Elite in the Governance of Higher Education. *Journal of Economic Issues*, *42*(4), 939–958. https://doi.org/10.1080/00213624.2008.11507197

[54] https://eric.ed.gov/?id=ED363233

[55] https://doi.org/10.1080/0260293980230207

[56] https://link.springer.com/article/10.1007/s10805-010-9117-9

[57] https://doi.org/10.1016/j.econedurev.2006.12.003

[58] https://doi.org/10.1080/00220670309596626

[59] https://doi.org/10.1016/j.stueduc.2016.08.007

[60] Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. Psychological Reports, 78, 1363-1372. https://doi.org/10.2466/pr0.1996.78.3c.1363

[61] https://scholar.harvard.edu/files/btl/files/bettinger_long_2010_does_cheaper_mean_better_-_impact_of_using_adjuncts_-_restat.pdf

private schools have historically been higher than their public counterparts. This is in part due to the selection of higher achieving students, but pre-college performance does not completely explain the difference. A 2010 study by Rojstaczer and Healy analyzed patterns within their database of over 160 colleges and universities from 1920 to 2006[62]. They found that on average private school students were graded 0.1 to 0.2 higher on a 4.0 scale for a given caliber of student (measured with SAT scores or a selectivity measure). When looking at grading across divisions, they found that on average science departments grade 0.4 lower on a 4.0 scale than humanities departments and 0.2 lower than social science departments. Evidently, both the institution and the division of a student's courses are correlated with a student's GPA.

Despite the increase in proportion of A's, this does not seem to reflect an increasing caliber of student. According to the 2019 National Assessment of Educational Progress High School Transcript Study, the average GPA has increased from 3.00 in 2009 to 3.11 in 2019 while over the same time period Grade 12 assessment scores decreased in mathematics and did not significantly change in the sciences[63]. Additionally, it is still debated as to how good of a predictor high school GPA is compared to standardized tests like the ACT and SAT. In a study comparing the predictive power of high school GPA against composite ACT scores, Noble and Sawyer (2004) found that across all levels of achievement, ACT scores provide greater differentiation than high school GPAs on success in the first year in college[64]. In particular, "at 93 percent of the institutions, a student with a 4.00 high school GPA had less than a 0.50 probability of earning a 3.75 or higher first-year GPA" in higher education and "in some cases, HSGPA values less than 3.00 provided little differentiation in terms of students' chances of achieving different first-year GPAs" (Noble and Sawyer 2004)[65]. On the other hand, some studies show that ACT scores and high school GPA are both valid predictors of first year performance (Westrick et al. 2015)[66]. Overall, there is insufficient evidence to suggest that the quality of students has risen, despite significant increases in GPA. GPA is no longer an effective tool for differentiating skill.

Falling Confidence in Higher Education

Grade inflation devalues education. The weight of a 4.0 GPA no longer carries the weight it once did. For colleges, participation in grade inflation lessens rigor, lowers quality of education, and degrades reputation (Chan, Hao, & Suen, 2007; Ehlers & Schwager, 2016)[67][68]. For

[62] https://www.gradeinflation.com/tcr2010grading.pdf
[63] https://www.nationsreportcard.gov/hstsreport/
[64] https://www.proquest.com/docview/225613390?sourcetype=Scholarly%20Journals
[65] https://www.proquest.com/scholarly-journals/is-high-school-gpa-better-than-admission-test/docview/225613390/se-2
[66] Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College Performance and Retention: A Meta-Analysis of the Predictive Validities of ACT® Scores, High School Grades, and SES. *Educational Assessment*, *20*(1), 23–45. https://doi.org/10.1080/10627197.2015.997614
[67] https://doi.org/10.1111/j.1468-2354.2007.00454.x
[68] https://econpapers.repec.org/RePEc:oup:cesifo:v:62:y:2016:i:3:p:506-521

students, they are deprived of feedback and left unmotivated by lack of recognition of exceptional effort (O'Halloran & Gordon, 2014)[69]. For society, grade inflation means graduates are unprepared for the workforce without the skills, dedication, knowledge, and work ethic desired by employers (Franz, 2010; Love & Kotchen, 2010; Yang & Yip, 2003)[70][71][72].

With more degree holders and higher GPAs, degrees are no longer a sufficient edge over other job seekers to obtain employment. Simultaneously, the rate of tuition increase has outpaced wages. This discrepancy has not gone unnoticed. Multiple news sources have published articles with headlines such as "Price Of College Increasing Almost 8 Times Faster Than Wages"[73]. This crisis has been expedited by rising housing costs and other costs of living. In order to afford degrees, federal student loan debt increased by over seven-fold between 1995 and 2017 (Durk and Perry 2020)[74].

Increasingly, the American public has been losing trust in higher education. According to a 2024 Gallup poll, reported confidence in higher education has fallen since 2015 from over 65% down to 36% in 2024[75]. Meanwhile the percentage of people reporting very little/no confidence has tripled from approximately 10% to 32%. Additionally, the gap in unemployment rates of Americans aged 25 and up by educational attainment has shrunk. What used to be a 5% difference in unemployment rates between those without a high school diploma and bachelor degree holders in 2005 is nearly halved in 2025[76]. However, the relative difference in median income of high school graduates and bachelor degree holders has stayed roughly the same since 2004, adjusted for inflation[77].

Flexible Grading in the 21st Century

In the early 21st century, research and usage of pass fail grading remained largely obscure. The most prominent use of pass fail grading was found in medical schools. Doctors are expected to be lifelong learners, staying up to date with the newest techniques, treatments, and health problems. To do so, the character of a doctor must be taken into account, particularly their intrinsic desire to learn. Additionally, higher than average rates of stress and burnout had been reported in medical students and the negative effects of distress have been well studied[78][79].

[69] https://link.springer.com/article/10.1007/s10734-014-9758-5
[70] https://doi.org/10.1016/j.econedurev.2009.10.013
[71] https://ideas.repec.org/a/pal/easeco/v36y2010i2p151-163.html
[72] https://www.asc.ohio-state.edu/yang.1041/grade-inflation.pdf
[73] https://www.forbes.com/sites/camilomaldonado/2018/07/24/price-of-college-increasing-almost-8-times-faster-than-wages/?utm_source=chatgpt.com
[74] https://eric.ed.gov/?id=ED610721
[75] https://news.gallup.com/poll/646880/confidence-higher-education-closely-divided.aspx
[76] https://www.bls.gov/charts/employment-situation/unemployment-rates-for-persons-25-years-and-older-by-educational-attainment.htm
[77] https://www.census.gov/library/stories/2025/09/education-and-income.html
[78] https://pubmed.ncbi.nlm.nih.gov/10926029/
[79] https://www.sciencedirect.com/science/article/pii/S0025619611610574

As a result, a wave of medical reforms were made including resident duty restrictions, self-development groups, and pass/fail grading systems.

In a 2011 review of pass/fail and well-being literature (1980-2010) in medical schools, it was found that all (four) of the papers reported improvement in some measure of well-being (stress, anxiety, depression, self-control, good health, level of satisfaction, group cohesion, and amount of free time) (Spring et. al)[80]. Student satisfaction was measured and found to have increased in two of the papers[81][82]. However, there were discrepancies in the long term effect of pass/fail with some claiming continued effect after the first semester while others found a return to typical levels in later semesters.

Spring et. al also reviewed an additional five papers (9 total) on pass/fail and academic outcomes (GPA, scores, residency attainment and performance). Grades were not found to be significantly different between pass/fail and tiered grading cohorts. Pass/fail cohort average significantly higher than the pass/fail cut-off in[83]. The pass/fail system was not found to adversely affect academic performance. However, acceptance into desired residency programs may be negatively impacted by pass/fail. In terms of residency attainment, roughly 73% of directors claimed they did not give preference to tier-grading schools and 33% of programs who filled all spots preferred students from tier-graded schools[84]. Surveys of students and directors also showed majority belief that pass/fail evaluation hindered the ability to compete for residency (29, 14). Although, other studies have shown that some students believe grades are already arbitrary while others prefer grading for motivation[85]. However, it appears that the actual impact of grading tiers may depend on the institution giving the grades as in practice when program directors were asked to compare Stanford's pass/fail classes with their own classes, only 3% of the Stanford graduates were judged as "poor" compared to their peer group[86]. There was also evidence that pass/fail grading reduced competition and external motivation for grades without decreasing the amount of time students spent studying, defying expectations of increased laziness[87].

Research supporting pass/fail and flexible grading continued to be published in the late 2010s. For instance, a decade long longitudinal study of medical student well-being found an 85% decrease in depression rate and a 75% decrease in anxiety in first-years when switching from a four-tier to two-tier grading system and restructuring their curriculum, among other changes (Slavin 2019)[88]. Additional studies on the differences between pass/fail and graded students

[80] https://doi.org/10.1111/j.1365-2923.2011.03989.x
[81] https://journals.lww.com/academicmedicine/fulltext/2009/05000/a_change_to_pass_fail_grading_in_the_first_two.28.aspx
[82] https://journals.lww.com/academicmedicine/abstract/1995/04000/the_effect_of_pass_fail_grading_and_weekly_quizzes.19.aspx
[83] https://pubmed.ncbi.nlm.nih.gov/7718068/
[84] https://pubmed.ncbi.nlm.nih.gov/7401142/
[85] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240587
[86] https://pubmed.ncbi.nlm.nih.gov/10353289/
[87] https://pubmed.ncbi.nlm.nih.gov/5548587/
[88] https://journals.lww.com/academicmedicine/fulltext/2019/06000/reflections_on_a_decade_leading_a_medical_student.27.aspx

found no consistent difference between student cohorts[89]. Interestingly, some have also suggested the usage of pass/fail as a method of combating grade inflation (Blum 2017)[90] while others argue pass/fail is a cause of grade inflation[91].

In summary, there has been an increased interest in pass/fail arising from a prevalent attitude that educational reforms were necessary. Researchers have found evidence that flexible grading systems reduce student distress, support collaboration, and encourage intrinsic learning without having a substantial impact on academic performance and test scores (White and Fantone 2009)[92]. However, there is also evidence that suggests that a two-tiered system makes it near impossible to distinguish between satisfactory and truly exceptional students, harming future career prospects as well as disincentivizing some students from putting forth their best effort and persevering through challenges. There is no definitive consensus in the literature about whether or not pass/fail is a better system than traditional A-F tiered grading.

Education During the COVID-19 Pandemic

Disaster preparedness, response, and relief is an important role that education fulfills. Schools help educate the public about how to prepare and act during disasters, and the return to school can serve to ease stress with its familiarity[93]. There is a plenitude of literature on crises such as school shootings[94], political conflicts[95][96], and natural disasters[97][98][99]. However, the COVID-19 pandemic was at an abrupt, unprecedented scale affecting the daily lives of nearly all communities.

On March 11, 2020 the World Health Organization declared a global pandemic. Policies such as travel restrictions, telehealth, social distancing, stay-at-home orders, and screening were implemented. In education, the response included remote learning, flexible deadlines, alternate assessment strategies, and relaxed grading policies. There is a growing body of research on the immediate and long-term effects of the pandemic on education. In the aftermath of widespread school closures in spring 2020 there is evidence of a negative effect of school

[89] https://europepmc.org/article/med/30392003
[90] https://core.ac.uk/download/pdf/132324597.pdf
[91] https://www.jstor.org/stable/27797025
[92] https://pubmed.ncbi.nlm.nih.gov/20012686/
[93] https://www.tandfonline.com/doi/abs/10.1080/02643944.2014.880123
[94] https://journals.sagepub.com/doi/full/10.3102/0162373715590683?casa_token=MLQZHaa7CC0AAAAA%3AnmMeAqWwjR5xoLHvw1Pf_HpCYkNPlN0WBcPF6sNBUXuPynJzsv3wYds0Wd4Ca5gFJZCQE66oPWRT
[95] https://academic.oup.com/jeea/article/17/5/1502/5292664
[96] https://journals.sagepub.com/doi/full/10.3102/0034654315609419?casa_token=S7N8UD6MVlMAAAAA%3AEAJI_7_3siHstlFQfw-laLYSxgUwJUEeVmI_yLS5VGCdTi743zFWVnaTrIhZZwoQfTy7srTJblOF
[97] https://www.sciencedirect.com/science/article/pii/S0890856709634675
[98] https://www.aeaweb.org/articles?id=10.1257/app.4.1.109
[99] https://www.researchgate.net/profile/Thomas-Devaney/publication/295702490_Impact_of_Hurricane_Katrina_on_the_Educational_System_in_Southeast_Louisiana_One_Year_Follow-Up/links/56cc9d5208ae059e37506abd/Impact-of-Hurricane-Katrina-on-the-Educational-System-in-Southeast-Louisiana-One-Year-Follow-Up.pdf

closures on student achievement[100]. In the three years spanning the pandemic (2020-2023), test scores were observed to have fallen compared to pre-pandemic levels and achievement gaps were amplified[101]. Numerous other studies support disproportionate impact of the pandemic on vulnerable groups[102][103][104]. Yet, grade inflation during this period exceeded multiannual trends and still persists in some academic divisions[105]. In addition to cases of COVID-19, students' physical well-being trended down as there was less engagement in physical activity[106]. Similarly, mental well-being also declined with increased world wide prevalence of depression and anxiety[107]. A plethora of research supports the idea that the pandemic exacerbated the pre-existing youth mental health crisis in facets from eating disorders to peer connectedness to substance use[108][109][110][111].

Intersection Between Flexible Grading and the Pandemic

The focus of this thesis is the impact of the pandemic on flexible grading policy. In the wake of school closures and remote learning, state DOE guidance often suggested or even required alternative grading[112]. A review paper on COVID-19 academic changes in higher education identified binary grading as one of their five key themes[113]. It's reported that at least 194 American universities implemented pass-fail policies during the Spring 2020 semester[114]. However, there is a dearth of research specifically evaluating the impact of COVID-19 induced flexible grading. In one study using data from Queens College suggested flexible grading policies helped reduce negative impacts of the pandemic, particularly for lower income students[115]. Another study using data during the first three semesters of the pandemic from a historically black college and university in North Carolina found that: utilization of flexible grading varied significantly between subjects, flexible grading was less likely to be used in general education courses, flexible grading use varies across socio-economic groups, freshman students are more likely to choose flexible grading, and STEM students were less likely to use flexible grading[116]. When looking at some of the most popular course sequences, they noticed a mix between

[100] https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.746289/full
[101] https://journals.sagepub.com/doi/abs/10.3102/0013189X221109178
[102] https://www.pnas.org/doi/full/10.1073/pnas.2020685118
[103] https://www.tandfonline.com/doi/full/10.1080/21681376.2022.2084447
[104] https://www.sciencedirect.com/science/article/pii/S0272775722000103?via%3Dihub
[105] https://www.tandfonline.com/doi/full/10.1080/03075079.2025.2470297?src=recsys
[106] https://jamanetwork.com/journals/jamapediatrics/fullarticle/2794075
[107] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02143-7/fulltext
[108] https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(21)00156-5/fulltext
[109] https://www.academicpedsjnl.net/article/S1876-2859(24)00021-4/fulltext
[110] https://www.mdpi.com/1660-4601/19/11/6768
[111] https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2786919
[112] https://eric.ed.gov/?id=EJ1374189
[113] https://www.tandfonline.com/doi/abs/10.1080/02602938.2022.2140780
[114] https://oudigitools.blogspot.com/2020/03/feedback-alternate-grading-in-crisis.html
[115] https://www.sciencedirect.com/science/article/pii/S0047272722000081?casa_token=vIy0DAM8D2MAAAAA:6xlm4a-PRCdeIUgQRzhnJWOnl69rgiO1T2vfUg9EdYaYK_2wpHSAdMIoGpbym3qbOVrkjH6Q
[116] https://pmc.ncbi.nlm.nih.gov/articles/PMC10199666/

students who benefited and those who were disadvantaged in the subsequent course. Perhaps related to this is a finding that in a study of two Canadian universities from the 2018-2019 to 2022-2023 academic year, in some subject areas GPAs have returned to pre-pandemic levels while in others GPAs remain inflated[117]. While there is no question flexible grading changed during the pandemic, its long term effects are unknown.

While a couple studies do examine the interaction between pandemic policies and flexible grading on student outcomes, their timeframes struggle to capture post-pandemic changes as the pandemic was not officially declared over until May 2023. While changes in behavior occurred in the height of the pandemic, it is unclear whether or not they persisted as education returned to pre-pandemic trends or if they will stabilize at a new norm. Additionally, both Rodríguez-Planas (2022) and Mostafa et. al (2023) used data from only moderately selective, affordable, and large public universities. There is a lack of research on impacts of pandemic policies on flexible grading on highly selective, prestigious, expensive, or private universities.

As students during the pandemic have begun and continue to graduate and enter the workforce, it is important to understand how pandemic policies have impacted them. For administrators, it is time to evaluate if changes in policies continue to be supportive in a post-pandemic world. With more disasters possible in the future, it's critical to understand the impact of the pandemic on settings including that of higher education. Hence, this thesis aims to determine whether or not there is evidence to suggest a change in student behavior under persisting pandemic changes. If so, I hope to identify patterns of change.

Data

All data used in this thesis is sourced from the Duke University Assessment Office through Jennifer Hill. Rows represent student-course pairs. In order to protect student privacy, data was masked and no socio-demographic data was provided. Furthermore, all data remains on in-office devices and I will no longer have access upon completion of my thesis. Please contact Jennifer Hill if you have questions/concerns about my base dataset.

Table TODO below shows a list of provided variables:

Variable	Type	Notes

While these data come from official records, there are some aspects that I remain skeptical of but assume are accurate for the sake of this analysis. Notably, it remains unclear to me

[117] https://www.tandfonline.com/doi/full/10.1080/03075079.2025.2470297?src=#d1e327

what the exact algorithm for calculating a student's academic level is. I experimented with matching a student's current semester number with their reported academic level. Results can be seen in Table TODO. While there are some significant discrepancies, I choose to believe the algorithm for academic level must include other factors such as progress towards degree completion. I also noted some questionable values for total enrollment in which I occasionally found classes claiming zero enrollment, yet that student still received a grade. After discussion with the Assessment Office, we believe it may be due to enrollment in a different section of that course being mistakenly used instead of totaling enrollment across sections. For instance, a course may be cross-listed in department A and department B but instruction is identical in both sections. While other potential accuracy issues may be present in my data, they should have relatively low impact on my overall analysis.

[insert num sem vs acad lvl bot table]

Methodology

Since my interest is in the impact of COVID-19 policy changes within the Trinity School of Arts and Sciences, only students who graduated from or intend to graduate from the Trinity School of Arts and Sciences will be included in this thesis. There may be a number of current students who later decide to transfer into Trinity or transfer into Pratt but I do not expect this number to be significant enough to be of concern. I established my baseline years as Fall 2012 to Fall 2019. Prior to Fall 2012, Duke used P/F grading in which students could only choose flexible grading after the declaration of their major (typically in sophomore year). Beginning in Fall 2012, Duke introduced the S/U grading system which, to my knowledge via the Duke Bulletins, is otherwise identical to the P/F system. I define the pandemic years as Spring 2020 until Summer 2 2022. The cutoff is chosen because by Fall 2022 there were no reported instances in which a period of mandatory remote learning was instituted (there were a couple days for which this was the case in Spring 2022) https://trinity.duke.edu/sites/trinity.duke.edu/files/documents/S_U%20policy%20history.pdf . While there may have been a couple odd remote classes that are only offered online (i.e. Cherokee) or exceptions for international students, it does not represent the experience of the vast majority of undergraduate students. Masking was still enforced in classrooms and the pandemic was only officially declared by WHO to be over in May 2023, but I claim its effect was relatively minimal and a return to normalcy was perceived. Data for post-pandemic semesters are Fall 2022 through Spring 2025.

Flexible grading policy has become more flexible in the aftermath of COVID-19, so I hypothesize that student utilization of flexible grading will be higher after the policy change when compared to the years prior. Specifically, students are now allowed to use S/U for general education requirements so I believe that any change in S/U will likely be reflected in general education courses. Curriculum 2000, under which all students in my data fall, general education requirements mandate that Trinity students to take two credits in each of five Areas of

Knowledge as well as two credits in each of five Modes of Inquiry. Unfortunately, I was unable to obtain these course codes as part of my dataset.

As a proxy for determining general education courses, I attempt to classify students by their academic plans under the assumption that most of the courses taken outside of the division(s) of their major(s) are taken for general education requirements. Trinity students are permitted to have up to three academic plans (majors, minors, and certificates) of which only two can be majors. Since minors and certificates tend to only require 5 or 6 courses and tend to be added towards the end, I do not classify students as belonging to divisions based on their minors and certificates. Due to privacy concerns and the potential to identify individuals given full records, course names, numbers, and departments were omitted. Instead, I was given catalog level (i.e. 100-199, 200-299, etc) and course division (Engineering, Arts & Humanities, Social Sciences, Natural Sciences, and Writing). In order to match majors with course divisions, I manually assigned each major with the appropriate division based on classifications at https://trinity.duke.edu/. While most majors were relatively straight-forward, I was unable to classify Program II and unlabeled interdepartmental (IDM) majors. These students are excluded from my model. A table showing classifications for each major can be found in the Appendix.

In addition to unclear majors, there are also a significant number of unknown majors in my data. Student academic plans in my dataset come from post-graduation records. This means that I lack any graduated students for the Class of 2026 (matriculated in Fall 2022) onward. Note that this is precisely the classes of students who fall after my pandemic time period (Spring 2020 – Summer 2 2022). I cannot afford to systematically have data loss during this period. To address this problem, I recognize potential bias but resorted to data imputation.

I used labeled data to inform my imputation, assuming that there has been no significant change in the distribution of courses a student takes in their major(s). Initial attempts included experimentation to determine the best threshold for number of courses and major divisions. However, using a number of courses does not extend well to students who have not completed a full eight semesters. In fact, it is plausible that some students remain undecided in their first several semesters, some students space out their requirements across all four years, and some students focus on their major and only later are reminded of general education requirements.

My current approach is to use proportion of 200 and above levels courses a student has taken per division to assign major divisions. With a 0.93%, 0.87%, and 0.87% accuracy for major divisions, the data suggested a threshold of 0.5, 0.4, 0.3 for Arts & Humanities, Natural Sciences, and Social Sciences respectively. The high threshold for Arts & Humanities logically makes sense, the more classes more likely to be majoring. However, the threshold for Social Sciences seems abnormally low as it insinuates a student who has a major in Social Sciences only takes roughly a third of their courses in Social Sciences. However, it should also be taken into consideration that 1) many students end up double majoring, 2) students might have minors or certificates in other divisions, 3) students could have other academic interests, 4) courses could be cross-listed across divisions, and 5) categorization of majors is somewhat debatable. To provide as example of 5, as defined by Trinity, African & African American Studies is classified as a Social Science while Asian & Middle Eastern Studies is a Arts &

Humanities. Looking at the confusion metrics for each proportion threshold in Figure TODO, it becomes evident that the increase in false positives with higher thresholds offsets the increase in true positives.

Another roadblock I faced is the possibility of S/U only and graded only courses in which students have no say in what grading method is used. Again, I lack this data and once again make assumptions. For classes that have 99% of all students receiving non-letter grades (i.e. S, U, withdrawal, etc) I assume they are S/U required courses like ECON 101. I believe this is reasonable in most cases as it is unlikely for all students to make the same decision, unless the class size is small. These mandatory courses are dropped from my model. However, I do not assume the converse is true. If all students in a course receive A-F (or withdrawal, etc) I choose not to assume the course is graded only. My belief is that in the majority of cases students would rather take a course on a graded basis as Trinity only allows 4.0 credits taken S/U to count towards graduation requirements. However, I do acknowledge that this means it is likely that my model will underestimate how likely a student is to take a course S/U in cases where the instructor prohibits S/U.

In addition to dropping the small fraction of courses that were unable to be matched (NAs for division, catalog level, and other course-specific attributes), I dropped courses belonging to the division of "Other". Broadly speaking, "Other" includes courses such as those for ROTC, Robertson Scholars, music lessons, and house courses. These courses are typically not taken purely for an "academic" purpose, and hence I do not consider them as qualifying a student to be in a true academic overload. In my opinion, half-credit music, dance, and PE courses also fall in a similar category as more of an extracurricular activity. I refrained from dropping all half-credit courses due to the existence of courses such as the half-credit STA 211 which was required for the major in Statistical Science. All full-credit courses in Arts & Humanities, Social Sciences, or Natural Sciences still remain in the dataset.

For modeling, my dataset has also been filtered to only contain courses taught online or at Duke's campus at Durham during either the Fall or Spring semester. This removes the impact of differences in environment, pace of instruction, grading institution, and other factors that confound these semesters and courses. Online courses are kept due to their prevalence during COVID-19 and their persistence for niche courses such those in the Cherokee Language Program.

In addition to constructing variables for major divisions and time period relative to COVID-19, I also engineered several other variables as I describe below:

- prev semGPA
- num plans
- took summer courses
- studied away
- term units

- num semesters
- something about overloads (total or so far….)

Results

Exploratory Data Analysis

Before fitting my model, I created several exploratory plots. In these plots, I use a more complete dataset than I do for my model by including mandatory S/U courses, non-traditional students, ….

*mention enrollement at Duke stable across years

[prop of grades given out A, B, etc by AY/term + GPAs]

[prop of overloads, S/U, # plans by CY]

[# S/U required(?) courses]

[prop S/U, S/U allowed courses, course sizes –> across divisions]

[acad lvl and catalog lvl]

[acad lvl/catalog lvl and S/U]

[# overloads and # S/U]

[time periods –> increase S/U decrease in W, 2014 vs 2019 vs 2024]

[student majors/divisions by AY]