

Incentivizing Higher Education Outcomes – The Next Frontier of Pay-for-Performance

Working Paper 4 in the Series: The Perils of Pay for Performance in Public Service Industries

[Burton Weisbrod](#)

Northwestern University and IPR

Version: April 8, 2025

Quotation and nonprofit distribution are permitted when accompanied by full author attribution and citation.

Abstract

Colleges and universities—both state institutions and private institutions—depend on revenue from students and their families (tuition), donors, public agencies, and grant sponsors for their economic health and existence. They compete for those funds based in large part on information concerning their performance, in education, research, sports, post-graduation employment, and alumni support—with expectations that have dramatically shifted and changed over time. Today, colleges and universities compete for rankings among a variety of published performance measures, and therein lies a significant problem. For strong financial rewards for *measured* performance encourages colleges to devote more resources to those forms of performance that are easily observed, measured, and highly rewarded, while other dimensions of performance are largely neglected. Even worse, it incentivizes “gaming” of those measures to enable the appearance of better outcomes. This paper examines the history of higher education institutions, their performance goals and expectations. It provides case studies of how performance measures tied to strong financial incentives have led in many ways to unintended and undesirable outcomes over time, and explains the underlying economic concept of why measurement itself leads to behavior changes. Finally, it makes a case for more nuanced and multi-dimensional measures of higher education performance, tied to weak rather than strong financial incentives, to better balance the goals and finances of educational institutions, their customers (students), and the public.

About the Series. This paper is the fourth in a series on “The Perils of Pay for Performance” for public service industries. The series highlights an important current issue, which is how for-profit firms, nonprofit organizations, and governmental agencies can coexist in many parts of a modern economy, with each playing a role supporting “better” performance. Other papers in the series delve deeper into those issues for other specific industries including K-12 schools and health care.

Acknowledgements. The research and discussions of performance measures for higher education was assisted by many people over the course of a decade of work. For this paper, I am particularly grateful for important editorial assistance and discussion refinement provided by Dr. Rachel Golden and Glen Weisbrod. I also thank the many Northwestern University undergraduate students who served as research assistants, specifically William Russell, Layne Kirshon, Andrew Ruth, Tyler Goff, Tory Do, Rui Du, Yushi Liu, and Robert Eli Winter. All conclusions and opinions are exclusively those of the author.

The Problem of Addressing Multiple and Shifting Goals for Higher Education

This is a tale of multiple goals and their compatibility with incentives. The four broad goals are to provide: (1) “Higher Education”, (2) For “All”, (3) Without “Breaking the Bank”, and (4) Avoiding Manipulative “Gaming” of “Quality.” But what do these goals mean, operationally? How do the incentives they provide affect their evolution, accomplishments, and failures? And most importantly, can the multiple goals all be achieved?

In the context of publicly supported industries such as higher educational institutions, the concept of Pay-for-Performance (P4P) refers to processes whereby the schools derive financial incentives and rewards from government agencies and public consumers as a result of performance outcomes regarding the quantity and/or quality of service they provide. This paper is the fourth in a series that examines how financial rewards tied to organization performance measures can actually undermine the achievements they are intended to encourage.¹ For higher education, part of this problem is that strong rewards for *measured* performance encourages colleges to devote more resources to those forms of performance that are easily observed, measured, and highly rewarded, while other dimensions of performance are largely neglected.

Compounding this issue is the fact that purpose and goals of colleges have morphed over time. Centuries ago, they were producers of luxury *consumer* goods serving only a tiny market niche among the landed aristocracy, and providing an essentially unchanging curriculum that contributed little or nothing to workers’ productivity and earnings. But now, a college education is increasingly justified as a valuable *investment* good that signals a worker’s increased human capital and is a near-necessity of modern life. How did we reach today’s heightened reliance on performance-based rewards as motivational magnets?

The goals of providing “higher education,” in such varied dimensions as subject content, class sizes, teacher experience, and other elements of quality, and making the evidence available to “all,” have become increasingly prominent in recent decades. But how to weight a college’s “success” in such varied dimensions as whether courses are taught online or in classrooms, with 20 students or 300, whether a student’s probability of graduating in, say, four years or six, is 95 percent or 3 percent, and whether the college is involved in financially “big-time” intercollegiate football or basketball, and is paying the football coach \$200,000 per year or \$4 million, are just a few elements of today’s concept of “higher education.” And the greater the rewards for a school’s performance, however it is measured, the stronger are the school’s incentives to game the incentive system by, for example, increasing class sizes or relaxing course grading standards to speed graduations. which cannot be easily translated into students’ lifetime earnings, donations to the college, or other sources of stakeholder support for the school, although student applications for admission may or may not be meaningful reflections. Attempts to game the measures of a school’s quality can bring profound effects on the road to the American Dream. It already has.

Higher Education for More Students, if Not for All: A Race of Quality to the Bottom?

The race is on. And more adverse effects lie ahead as increased social pressure on state access to college causes more schools to curb their costs, with uncertain effects on educational quality. Greater

¹ Earlier papers in this series are: Weisbrod, “Why Strong Performance Rewards...”, Weisbrod, “The K-12 Education Revolution...”, Weisbrod, “Financial Rewards for Health Care...”

overall access to college does not specify the mix of students among racial or cultural groups, nor to the fields of study that will be available, class sizes, and various elements of instructional “quality”.

People are now living longer, working later in life, and in the process confronting changes that are not only restructuring labor markets but in ways that are more difficult to anticipate. Both men and women are finding that the education and training they received in their youth are inadequate for the changing job labor markets decades later. For the first time in American history, more women than men are now graduating from four-year colleges, entering the labor market, and learning, perhaps much later, that their college education was not suitable for their unpredictably changing interests and opportunities. The now-booming market in AI –Artificial Intelligence – is bringing vast, but still largely unknown, changes in labor markets for today’s college students during the coming decades of their working lives.

The lifetime entwinements of college education and shifting labor markets are bringing mounting pressures on government – many focusing on cost containment such as tax deductibility of charitable donations to nonprofit colleges, property-tax exemptions for those schools, subsidies and “forgiveness” of student loans, and other fiscal instruments for expanding college attendance.

Increased access to higher education, if not for “all” then at least for a substantially increased portion of the population, took a major leap forward in 1944, toward the close of WWII, with the first in a series of Federal *G.I. Bill of Rights Acts*. Military veterans received financial inducements to complete college degree programs, presumably expecting to raise their future earnings. But even as the G.I. Bills associated with the Korean and Vietnam Wars, in 1950-53 and 1955-75, ended, the fiscal goal of avoiding *breaking-the-bank* was not out of sight.

Neither is it today. Cost considerations remain central to the process of expanding access to post-secondary, *higher*, education. So do concerns about the matching of college programs with the uncertain future composition of the job-market in the decades ahead. Government payments to a college, or to students who had received government loans to help defray tuition, fees, and other education costs that cannot be evaded even through bankruptcy, leave student borrowers vulnerable to overly optimistic forecasts of future earning power that may be inadequate to repay loans and interest.

Indeed, schools that are eager to attract greater income from tuition, fees, and other student-related sources have the incentive to game prospective students who can be convinced to attend the school and borrow the necessary funds, underplaying the uncertainties of their long-term ability to repay the loans. The need to establish efficient incentives highlights the importance of college finance arrangements and revising them as new gaming mechanisms evolve. The 2022 education loan-forgiveness legislation is the latest recognition of the forces competing to attract more low-income students, but make loans even more burdensome if borrowers’ post-college earnings prove to be insufficient for the loan repayments.

Financial incentives have also been taking their toll through their effects on higher education quality. Even after centuries of production stability since the industrial revolution, in pre-British colonial days, the history of higher education remained a story of little change in what colleges did, who they served, and how their goal achievements were measured and rewarded, by students, their parents, governments, and private donors. Yet it was also a tale of perils -- of using strong rewards for poorly measured performance.

Higher education is the industry next in line, after k-12 schooling and health care, to turn to stronger incentives to measure and reward specific dimensions of performance. Higher education has long ago

been transformed from a labor-intensive production process that was relevant to only a tiny minority of youth from wealthy families with little or no concern about effects on future earnings, loans, or their repayments. Now, however, in the 25th century, being a college graduate is not only increasingly commonplace but is seemingly essential to even modest opportunity for greater earnings.

The role of higher education has been revolutionized since WWII. In 1944, with the *end* of WWII in sight, the enactment of the first in a series of Federal “G.I. Bills of Rights,” brought a massive increase in support for higher education as well as for housing and other benefits to literally millions of honorably discharged military veterans. The 4.6 percent of the adult population aged 25 or more who were college graduates in 1940 vaulted eight-fold, to 37.9 percent by 2021.²

The rarity of being only a high school graduate soon became history. So did having just “*some*” college schooling, and, increasingly, even being a college graduate. Now, having at least some college education has become so commonplace that it is almost useless as a signal to employers of employee skills and expected labor market productivity. So new screening mechanisms were needed beyond such traditional higher education performance measures as college degrees and impressive Grade-Point Averages (GPAs). But can higher education continue to serve this sorting function?

Yes, it already is. But what the measures convey about future labor-market performance of workers who attain just “*some*” college education, or even a bachelor’s degree, are of declining relevance. It is no accident that as becoming a college graduate has become more *common*, so it has become less *useful* as a sorting mechanism for signaling greater worker productivity. When almost every worker meets some particular performance test, in higher education or elsewhere, that mechanism loses its information value for conveying workers’ expected future productivity. The solution? Develop and promote another way to distinguish among potential employees -- why not encourage post-baccalaureate training by expanding Master’s Degree programs? That, too, is happening.

In fact, the number of awards of Master’s Degrees has been escalating for decades. Back in 1970, the number of Master’s Degrees awarded in the U.S. was 27 percent of the number of Bachelor’s Degrees awarded that year, but by 2022 it reached 44 percent.³ And the narrowing of the growth gap between degrees continues, most rapidly in teacher education and business administration, but why the growth of Masters Degrees, and why in those fields?

The increase in Master’s Degrees being awarded, compared with Bachelor’s Degrees, is a signal to employers that job candidates with more advanced degrees have greater motivation and ability to learn, and greater expected on-the-job productivity. To be sure, though, neither degree connotes a homogeneous labor-force population, although the information that a new labor-force entrant holds a Master’s degree, rather than just a Bachelor’s degree, may be useful not only for employers but for other “matching” processes such as finding marital partners.

There are downsides, of course, to the use of higher levels of academic degrees and certificates as information about workers’ expected labor market performance: (a) A Master’s Degree, as a mechanism for *sorting* job candidates, is *costly*, varying among colleges and fields of study, but ranging between \$30,000 and \$120,000; (b) *financing* those costs often lead students to take out large loans that cannot be erased even through bankruptcy, nor repaid because their post-degree earnings are insufficient; (c) *competition* among colleges for tuition-paying students can be expected to encourage debasement of

² U.S. Census Bureau, Current Population Survey, Educational Attainment...

³ (NCES) National Center for Education Statistics, Digest of Educational Statistics...

education quality in ways that are often not easily recognized; for example, colleges are replacing lifetime, tenure-track, faculty, with temporary, adjunct, faculty, and expanding offerings of non-traditional college courses such as “beer brewing, which” may advance a Master’s Degree program but be short-lived.

Other Performance Metrics and Gaming in Higher Education

A 2012 report on *Improving Measurement of Productivity in Higher Education* by the National Research Council of the National Academies recognized the clashes among *measuring* performance, its durability, and its effects on *incentives* to *game* a school’s measures of its performance. The Report’s Appendix on “Commonly Used Performance Metrics for Higher Education” included such measures as a school’s Graduation Rate and its Student-Faculty Ratio.⁴

These measures are used commonly for two prime reasons – they are seemingly easy to measure and to understand. Their side effects, however, are not as clear-cut; being in common usage does not imply that they cannot be gamed to evade penalties or induce unintended rewards. But to say that certain measures of performance are *commonly used* also does not mean that they have no *systematic* shortcomings; they do. The National Research Council had captured the essential point, citing sociologist William Bruce Cameron: “Not everything that counts can be counted, and not everything that can be counted counts.”⁵

True, but with a caveat. In the context of public and nonprofit sector activities in general, and higher education in particular, aligning *pay* with *performance* is more than an exercise in *counting*; *each* component of performance must be both *measured* and *valued* – accurately and without systematic over- or under-counting or valuations. Otherwise, stakeholders face incentives to *over*-provide types of performance that are easy to exaggerate – to over-count and/or to over-value – while *under*providing types that can easily be under-counted and under-valued.

Sociologist Cameron’s insight remained valid – but it also remained incomplete and misleading. Whatever the industry involved, rewarding what can be measured *easily* is a path to inefficiency, because of what it *fails* to count and value accurately. Yet this is the path linking rewards to performance on which public policy has embarked; *measuring* and *rewarding* forms of performance when that is feasible at low cost but ignoring them when they are costly to measure, and reward is a recipe for misallocations and inefficiency.

So, it is not surprising that the most commonly used performance measures in higher education pay little attention to hard-to-value output quality, but considerable attention to easily-measured attributes. Considering the National Research Council overview of performance metrics and their consequences; the emphasis on *ease of measurement and valuation* is clear.⁶

As a society we know how to *produce* a “more educated” population – at least as that is measured by years of schooling, degrees awarded, and student Grade Point Averages (GPA). We also know how to *measure* the *costs* of producing the results – in total, per student, and per degree awarded. But there are hurdles, which are elusive and changing: how to *define*, *measure*, and *reward* the various elements of *quality*, and how to decide which are “acceptable” in the interest of cost control, such as by

⁴National Research Council, *Improving Measurement....* p. 137-143.

⁵ National Research Council, *Improving Measurement....* p. 10.

⁶ National Research Council, *Improving Measurement....*

increasing “class sizes.” And many elements of quality can be gamed by teachers and administrators who can choose, for example, whether to count “students” who come to class “late” or are absent “often.”

Some college performance metrics such as student GPAs and graduation rates have appeal because the data for calculating them are readily available to the school and familiar to its students and their parents. Yet there is a downside -- the incentive for colleges to game the *grading* system, allowing GPAs to rise, graduation standards to be reduced, cutting the number of course credits required for graduation, and making it easier for a student to accumulate credits for prior “life-experience” which makes graduation faster and cheaper.

The critical problem with performance measures is not that they are imperfect – of course they are. The problem is that the measures are easily *biased* through intentional gaming; costs are systematically *understated* and accomplishments *overstated*. Both result from our limited ability to measure true “performance” accurately, and so to reward it, as well as from the all-too-common situation in which performance is reported by the very stakeholders who will benefit from *reporting* lower costs and higher performance benefits. By doing our “best” to measure and reward “costs” and “benefits” we end-up measuring what is *easily* measured, and in ways that have favorable effects for the person providing the information, albeit distorted. The distorted performance data erodes the rationale for measuring it.

So, the process of *gaming performance* is predictable, understandable, and counterproductive. It is especially problematic in the public and nonprofit sectors of the economy, because better performance, as it is measured, is typically less costly than truly better performance, as we saw earlier in this series of papers -- in a car manufacturer’s manipulations of its reports to government air pollution regulators on their cars’ pollution emissions,⁷ on K-12 schools’ deceptive scoring of standardized tests by changing students’ wrong answers to right ones, to “improve” their test scores,⁸ and on Veterans Administration Hospitals’ under-reporting the weeks and months of patients’ waiting to see a physician⁹.

Gaming of evidence on performance evidence, to qualify for greater rewards, is deeply troubling because it is so easily overlooked. Evaluations of regulatory programs and the efficiency of free-market competition typically build on the premise that the reported data, regardless of their source, are *accurate* or, at least *unbiased*, estimates of the truth. But they often are not.

The common retort that “no data are perfect,” while true, misses the point. Perfection, indeed, is not worth pursuing, generally because of its cost; and if the measurement imperfections are not serious, especially when they are *not* systematically biased upward or downward from the truth, the random errors of measurement may have little effect, although that cannot be assured.

The conflict between the self-interest of stakeholders who report their own performance, which induces upward bias in their performance reports, and their responsibilities to the organizations to which they report, are clear.

- Only the easily measured dimensions of performance will be reported, because of the low cost. But if the reporting of performance is done by someone who is personally rewarded for “better”

⁷ Weisbrod, Burton, “Why Strong Performance Rewards...”

⁸ Weisbrod, Burton, “The K-12 Education Revolution...”

⁹ Weisbrod, Burton, “Financial Rewards for Health Care...”

performance, or if downwardly biased reports result in penalties, dimensions of performance that are particularly costly to measure without bias will be under-rewarded or under-penalized.

- Whichever outcomes are *systematically* under-measured or over-measured, and under-or over-rewarded, will receive either too little or too much attention, rewards, and penalties compared with the efficient levels that their performance justifies. These are the social welfare losses that gaming generates.

Grade Inflation: More than One Way to Game Teacher Incentives

Grade inflation is not only a theoretic possibility; it is a predictable response to incentives, for colleges, their faculty, students, and, later, employers. In the late 1930s a college course grade of “A” or “A-” was a significant achievement, constituting only 15 percent of all U. S. college and university grades. And 35 percent of grades were in the C+ to C- range.

No longer. Some 80 years later, there had been a tidal change in the distribution of grades. The share of high grades, “A” and A-, tripled, from 15 percent to 44 percent, while C range grades plummeted from 35 percent to 15 percent, although the changing patterns had not been constant over time.¹⁰ Inflating grades has been an inexpensive way to convey to prospective college students and their parents that the school was performing “well.”

Another example occurred in 2023, when two deans of Harvard College brought critical attention to the prominence of “A” grades, which had increased from 60% percent in 2011 to 79% in 2021.¹¹

Similar incentives apply to a college or university president, or to a football or basketball coach, if the contract included a bonus for a particular measure of performance. For instance, one study found that when a Big Ten Conference football coach’s contract specified a bonus in addition to base salary, there was a maximum bonus for players’ *academic* performance (that is, Grade Point Average), which was less than 10% of the maximum bonus for winning a specified number of games in a single season. The message and incentive for the coach were clear: concentrate on recruiting star athletes rather than on players who were good enough to make the team but were academically more impressive.¹²

Performance rewards that incentivize football coaches to concentrate on winning games and championships remain widespread. But other studies of coaches’ contracts have shown the persistence of winning games as a goal, rather than advancing players’ academic skills.¹³

Gaming a College’s Performance: Self-Reporting its Alumni Giving Ratio (AGR)

A school’s reporting its own performance is an invitation to gaming – to circumventing, if not ignoring, external rules, regulations, and methods for achieving a better school.

Example 1: Oklahoma. *U.S. News & World Report (USNWR)* publishes its *Best Colleges* survey, the country’s leading college-rating organization. The performance data on the proportion of the school’s living alumni that donate financially to the school is one determinant of the school’s national quality ranking by *USNWR*. A higher Alumni Giving Ratio (AGR) was assumed to reflect greater alumni satisfaction and thereby, presumably, higher school quality.

¹⁰ Rojstaczer and Healy, *Where A is Ordinary*....

¹¹ Hamid and Schisgall, *Harvard Report Shows*...

¹² Weisbrod, Ballou, and Asch, *Mission and Money*... p.255.

¹³ Weisbrod, Ballou, and Asch, *Mission and Money*... 255.

As with many types of information used to rate or rank institutions such as colleges, universities, museums, hospitals, and charities, the cost of obtaining *accurate* information about consumer satisfaction can be substantial, but the *lowest-cost* source is often the organization itself. So, it is not surprising that *USNWR* used alumni-giving data provided by each school for itself, nor that each school faced a conflict-of-interest, a “double-agent” dilemma between the incentive to raise its national rank and to report data honestly and accurately.

In 2019 it was reported that the University of Oklahoma had allegedly been deliberately overstating its AGR – the percent of its living and reachable alumni who had donated to the University for each of the previous 20 years, since 1999. The University had recently informed *USNWR*, however, that its latest AGR, which it had reported as 14 percent, was exaggerated and should have been only 9.7 percent.¹⁴ But the causes of the overstatements were not made public.

USNWR was and remains a private for-profit firm that uses college-supplied data in its calculations of college rankings, then disseminating them widely to prospective college students, their parents, and other stakeholders. A school’s ranking could be intentionally falsified, and *USNWR* was intent on protecting its property rights by penalizing schools for over-stating their AGRs as measures of alumni satisfaction. The magazine penalized Oklahoma for its overstated reports -- dropping the school entirely from the 2019 ranking, because of its *admitted* 20 years of overstated reports of the number of its alumni who donated to the University and were presumably satisfied with the school.¹⁵

The facts that *USNWR* was and still is a private firm meant that the University was reporting on its *own* performance, made it likely to be self-serving and its accuracy suspect. While the governmental University of Oklahoma, which had admitted its data falsifications, was penalized by being dropped for a year from the *national* ranking; the *for-profit* publisher, *USNWR*, was not penalized, unless bad publicity cut its sales and profit. And the University was not unique.

Example 2: A Liberal Arts College in Michigan. A small nonprofit college, Albion, with about 1,250 students, developed another way to game the *USNWR* ranking process and its alumni appraisals of the school. It developed a devious way to boost its reported Alumni Giving Ratio, and in the process to elevate its “quality” ranking. It was a success.

Albion’s President had learned that a small but meaningful five percent of a school’s *USNWR* total score, which determined its national ranking, was the school’s Alumni Giving Ratio (AGR). So the challenge for Albion was to find a low-cost way to increase its self-reported AGR, which would almost certainly raise its national ranking.

Albion developed a new way to report – though falsely – its “quality,” as measured by its AGR. The technique manipulated the calculation of its own AGR and then reported the inflated performance data to *USNWR*.¹⁶ The AGR was simple in concept, easy to calculate, and useful for a school’s appraisal by its alumni; it was also easily gamed to elevate the school’s ranking -- if other schools did not follow suit.

Albion’s Administration found a way to increase its reported AGR. To see how, consider the *USNWR* definition of the AGR: First, the measure did *not* depend on the total donations it received from alumni in a particular year -- only on the *percentage* of alumni who donated. But the AGR *did* depend on the

¹⁴ Gifford, Univ. of Oklahoma fudged...

¹⁵ Gifford, Univ. of Oklahoma fudged...

¹⁶ Golden, Math Lessons to Boost...; Weisbrod, Ballou, and Asch, Mission and Money...

number of its alumni donors as a percentage of all its alumni for whom the school had “good contact data” – but how was that to be measured?

So both the numerator and denominator of the AGR donation ratio could be gamed by a school wanting to show its alleged success in satisfying its former students. Albion did that, and effectively.

First it dichotomized its alumni into those who did and did not make a donation in a particular year; the larger the share of its reachable alumni who expressed their satisfaction with the college by donating to it that year, the more the school would rise in the national ranking. Albion capitalized on that, but a crucial question emerged; how *should* an alum who had *not* donated for some arbitrary number of years be counted – especially if the school was uncertain whether the alum was still alive and so should be counted in the AGR denominator? How much time and expense should a school be expected to devote to determining whether a non-donor alum was alive? Are those standards enforceable?

USNWR provided schools with no answers. Yet a school did have to submit AGR donor information to be ranked, so *USNWR* had to decide how to set rules and enforce them.

From an efficiency perspective, what should a college be incentivized to do before submitting its AGR information? Given the benefits a college can expect from being ranked higher, and the costs of its achieving a higher rank by *truly* performing better rather than by using some form of statistical gaming—legal or not -- and rather than increase or decrease efforts to locate *non*-donating alumni, could it be cost-effective for a school to simply *assume* that some alumni who did not donate in a given year were “unreachable.” That gaming would almost certainly be legal and would elevate the school’s national ranking.

More General Ways to Achieve High Alumni Giving Ratios. Note that if a college reached a judgement that at least some of its non-donating alumni were either no longer alive or were unreachable, that would *not* alter the *numerator* of the school’s AGR statistic, so the number of its *reported alumni* donors would not be changed. But re-classifying some *non*-donors as being either *unreachable* or *deceased* -- the AGR *denominator* -- would no longer count alumni who did not donate. With the AGR numerator unchanged while the denominator decreased, the AGR ratio would rise, as would the school’s expected rank.

Albion College found a second gaming strategy to raise its ranking. Again, there were two routes, both essentially costless to the college but imposing costs on other stakeholders:

- 1: Albion asked, but did not mandate, each of its graduating seniors, who were about to become alumni, to make a one-time donation of \$30. So far, so good; then came the gaming.
- 2: The college *recorded* and reported each \$30 donation not as a single contribution but as a \$6 contribution for each of the following five years. The result of this accounting trick was that a single \$30 donation was reported to *USNWR* as an alumni donation for each of the next five years, so the contribution elevated the school’s Alumni Giving Rate in six years, whether or not the alum gave another donation in the other five years.

Why the initial request was for only a \$30 contribution, and why each was then arbitrarily reported over only five years rather than for a longer period, was also not reported. The *effect*, though, was clear; a single donation to the school increased its reported AGR not for just one year but for six, likely elevating the College’s *quality* ranking each year.

The underlying problem was one of *mechanism design* -- of harmonizing the performance measurement process with the goals and incentives of multiple stakeholders. In this case the tension was between the interests of potential student applicants who sought information on *prior* students' satisfaction, and the interests of colleges competing for students; a school that was more adept at gaming the information available to prospective students would be more successful in attracting students than schools less skilled at statistical gaming. In short, schools had the incentive to develop and self-report data in ways that keyed-into their informational advantage over both prospective students and the schools with which they competed.

Figure 4.1 sketches a number of programs in higher education that I discussed earlier in this paper, highlighting examples of the stakeholders and legality issues that may apply to them. This table has some similarities and differences compared with parallel actions in other industries covered in prior papers in this series, on hospitals, nursing homes, and K-12 schools.¹⁷

Figure 4.1 Forms and Gaming of an Alumni Giving Ratio (AGR) in Higher Education

Legality of the Gaming	Stakeholders			
	College Officials	Prospective Students	Alumni	Other Colleges
(A) Illegal	AGR	AGR	AGR	AGR
(B) Legal but misleading	DS	AGR	DS	DS
(C) Unclear legality	BAO	?	?	BAO

AGR manipulating measurement of a college's Alumni Giving Rate (See text, above).

DS Donation Spreading -- Reporting a donation in one year as five years of donations of one-fifth of the contribution (See text, above).

BAO Bribing a college admission officer to admit a student from a wealthy family.

"Double-Agent" Problems Associated with Self-Reporting

Fundamental to many performance-measurement issues is a dilemma: obtaining *information on performance* generally conflicts with obtaining information at low cost and with confidence that it is accurate and unbiased. Important performance information is often knowable, but is known by a stakeholder who does not have the *incentive* to reveal it in an unbiased way -- neither exaggerated nor underestimated -- to another stakeholder such as a government agent, a parental payer of tuition, or a college fundraising ("development") officer.

The underlying problem is that in many circumstances a better-informed stakeholder is dealing with a less-informed stakeholder, and in such asymmetric information cases, the better-informed party has the incentive to selectively give, withhold, or manipulate the information that is transmitted to other stakeholders. Better-informed stakeholders have the incentive to reveal, selectively, only information

¹⁷ Weisbrod, "Why Strong Performance...", Weisbrod, "The K-12 Education...", Weisbrod, "Financial Rewards for Health Care..."

that would advance his or her own well-being. The result of informational asymmetry is predictable and socially inefficient, as George Akerlof, a Nobel Laureate in Economics, showed in 1970.¹⁸

The gaming of a P4P reward-system and its biased reporting of *performance* are often encouraged, unintentionally, by the reporting system. The reason: a stakeholder's performance and the associated rewards are often self-reported by the very stakeholder whose performance is being judged and rewarded -- be it a college, faculty member, student, a government tax collector or tuition-payer, or a commercial college-rating organization such as *U.S News & World Report* (USNWR). The result: when, for example, a college employee (1) provides information to another stakeholder, say USNWR, and if (2) the employee's compensation from the college depends, even partly, on improvement in the school's ranking, the employee is thrust into the position of a "double agent." Providing upward-biased performance information might well benefit the college and him, personally, but at the expense of students who rely on upward-biased "performance" information from the school's reports on itself.

Because a college is often a double or even triple-agent -- trying to advance its own self-interest, reputation, and revenue, while also serving the educational interests of its students and the public policy goals of government, it has conflicting incentives. Casualties result.

Rewarding educational "quality" in a climate of change

Stronger P4P incentives are generally not intended to depress quality, but they typically do, because of the costs of *observing* education "quality," *measuring* it accurately, and *valuing* it. The results: forms of quality that are easily and accurately measured are well rewarded, others are not.

Within this information-cost framework, today's college faculties have become less costly to access and reward, as part-time and contingent faculty are substituted for the historically higher-cost faculty with longer-term employment contracts. Better-informed college administrators can game the less-informed students and their parents; the full effects of cost-cutting practices on the quality of higher education hinge on the costs of measuring their long-term effects, even when their short-run presence is easily observed, as with class size and course grades.

Higher education, in short, is now on a historically unprecedented path, in many ways analogously to health care, to expand access to more of the population and to a widening range of more costly medications, diagnostic tests, and surgical treatments. The goals in both health care and education are lofty, but also more diverse, cost-increasing, and more open to gaming.

These are not just possibilities; they are already visible. Low-enrollment programs and courses are being terminated, class sizes are increasing, and program quality is giving way to the search for new ways to generate revenue and reduce costs. Enrollments in Master's Degree programs are also increasing. As the share of the labor force consisting of college graduates has risen, and as that once-rare status has become commonplace, a growing share of graduates has sought ways to distinguish themselves from other baccalaureate degree holders, and the number of holders of Master's Degrees will continue its upward trajectory; the days of college graduates being an elite group of children of the super wealthy is bygone.

¹⁸ Akerlof, The Market for Lemons...

Gaming by College Faculty: Research and Teaching Tradeoffs

It was back in 2013 when the provost of the University of Vermont set out to strengthen the school's programs and improve its ability to compete. He focused on increasing faculty scholarly productivity and on ways to measure it, asking deans and departments to develop their own performance measures.

"Some [of the university's] colleges, including the business school, created a tiered ranking of journals – a faculty member would receive eight points for being published in a top-ranked journal, down to one point for a 'fourth tier' journal. Earn 24 points – three publications in a top-tier journal – and see your teaching expectation for the year drop from five to three courses. Earn fewer than six points, however, and you might get a sixth course to teach."¹⁹ That was a clear Pay-for-Performance system.

Yet the penalty to faculty for low productivity in research was surprising: penalize the offending faculty by demanding that they *teach more* -- in effect, by penalizing their students. Yet another example of the difficulty of aligning rewards (or penalties). But what evidence is there that an unproductive researcher can be expected to be more productive as a teacher?

Much depends on the definition of "productive." But more: there are two elements of productive performance – measurement and valuation. If research is measured in number of publications, for example, while performance in teaching is gauged by the number of courses or students taught, shifting a faculty member's allocation of time from research to teaching would have unknown overall effects on performance over time, even if short-term research productivity was already at zero; the length of the period of research accounting would have complex long-term effects.

But what, exactly, is a "publication?" Should its length matter? How should an article with multiple authors be counted? What is a "top-tiered" journal, and should an article published in a lower-ranked journal be regarded as equivalent to non-publication? How should a faculty member's "creativity" be measured? How can a faculty member whose promotion, salary, bonus, or other rewards or penalties be expected to respond to incentives?

Of course, the greater the rewards and penalties are, the more powerful are the likely perils, no matter the degree of high-minded intentions to encourage social efficiency. So the extent of dangerous consequences of manipulative gaming inevitably has two elements –how accurately performance is measured, and how accurately the measures are valued.

The central point is not simply that incentives matter, but that powerful incentives can be counterproductive. In higher education, key personnel, in this case faculty, can take advantage of systematically biased measurements and valuations – particularly when performance data are reported by the very stakeholders whose performance is being gauged.

If an additional faculty publication would garner a reward for each author, a clear incentive would be to encourage multiple authors -- or not, depending on how much, if at all, the addition of a co-author would reduce the rewards to other authors; If the productivity-reporting and reward system gave each author full credit for the publication, the incentive to add co-authors would certainly be greater, inducing an upsurge in multi-authored papers, compared with a reward system that allocated a total of

¹⁹ Patel, How One University...

one article-credit per publication, shared among the authors. That performance-reward approach would encourage less collaboration and fewer co-authored papers.

***Other Ways Faculty Can Game their Own Performance Rating:
Publications, their Lengths, Citation Frequency, and Prestige***

If faculty performance is gauged not simply by number of publications but by the number of times each one is cited in another paper, the resulting incentive is for authors to cite the publications of their friends and colleagues: increase the rewards for “better” performance – higher citation-counts -- and then watch the intended incentive effects emerge.

Back in 1947, economist George Stigler had recognized the attractiveness – as well as the dangers -- of a *merit*-based P4P reward system for the faculty at his fictional South American University. His school *adopted rewards for better faculty performance, as measured by the number of his or her publications*, perhaps weighted by their length, frequency of being cited by other authors, using measures that were easy to understand, tally, and reward.

Stigler also saw the *perils* of *strong* incentives – the counterproductive behavior that would be encouraged, unintentionally and undesirably, if faculty were paid on the basis of the number of his or her publications. The system could be gamed, of course, yielding larger rewards but without real increases in scholarly performance. That, indeed, was what Stigler wanted the readers of his “proposal” to recognize.

It would not be difficult for faculty to arrange deals with colleagues to their mutual advantage but not to the school’s; they could take on co-authors, thereby increasing each co-author’s measured productivity and compensation. True, for every gaming mechanism a way could be devised to penalize its “abuse”; there would be costs, though, of monitoring and enforcing rules that either mandated or prohibited illegal particular actions deemed to be *improper or unintended*. Yet each such regulatory restriction or mandate provides new opportunities for ways to game the reward and penalty system.

At first the publication incentives in Stigler’s fictional university “worked” – *sort of*. They encouraged faculty to increase their productivity, as it was being measured. Faculty recognized that each book they were writing would count for but one publication. The obvious incentive was to stop writing books, instead converting each chapter into a publishable journal article. One book already in page proof, was withdrawn, each of its 19 chapters was published as an article. And a faculty member who received an offer from another university could be given any number of performance points the department chairman “deemed fit.”²⁰ The subtext was clear: whatever the details of the rewards for performance, faculty and administrators could and would engage in counterproductive strategic gaming.

The rector at Stigler’s fictional university set out to increase faculty productivity, by developing a merit reward system measuring and rewarding each faculty member’s “performance,” thereby aligning the private interest of a faculty member with the collective interest of the university. An elaborate system of competitive faculty exams was developed, permitting, for example, junior faculty to replace senior faculty by scoring higher on the test. The Dean expected the system of more powerful rewards to strengthen the university and its reputation.²¹

²⁰ Stigler, *An Academic Episode...*

²¹ Stigler, *An Academic Episode...*

Not so fast. When faculty discovered that they would be rewarded for publishing more journal articles but not books, they saw the message, however unintended it may have been, was to discontinue book-writing projects and publish individual chapters. That was just the beginning; when faculty discovered that lengthier articles were being rewarded more, they shifted submissions to journals that used wider margins and so printed fewer words per page, resulting in a paper of a specific number of words being longer in journal pages, and so, more handsomely rewarded. Faculty ingenuity in finding “loopholes” in the system of measuring their performance provided incentives for them to game the reward system.²²

Anticipating and suppressing gaming

No matter how higher education’s performance is gauged and rewarded, whether classes are larger or smaller, whether they are held in traditional classrooms or online, and whether teaching is by faculty that are full-time or part-time, lifetime-tenured or short-termed adjuncts, the circumstances will translate into incentives to game the reward system. New avenues will emerge for colleges, students, and donors -- private and governmental -- to respond to new incentives and opportunities, be they legal, illegal, or borderline.

Colleges know, for example, how many of their courses are being taught by temporary or part-time faculty, even if students and parents do not know, or even care, and whether greater use of adjunct faculty affects its “quality,” let alone how to value that.

There are takeaways, of course. As the goals for higher education take on more sweeping forms, as ways of measuring success expand, and as opportunities for stakeholders to game the performance reward system increase with the number and complexity of success measures, the adverse side-effects of *stronger* rewards for poorly measured performance *will grow*. Effects on college quality, seemingly easily observed indicators of a college’s performance, such as a more diverse student body, slowing the escalation of tuition, expansion of athletics and other extra-curricular activities, as well as divestment of school endowments from fossil-fuel investments and other environment-degrading air and water products, will present more opportunities to game the measures of performance. They already have.

The Historical Context for Viewing the Past and Future of Pay-for-Performance in Higher Education

The future of collegiate pay-for-performance in higher education has increased gaming and deception, *but it is* not a historic relic fading away over time. To the contrary, it is gaining momentum that will continue to increase, responding to incentives that encourage devoting time and effort to forms of performance that are easily measured and well-rewarded, while weak incentives to improve performance are associated with activities that are difficult to measure and poorly rewarded. Yet higher education is being called-upon to operate in both arenas -- accepting responsibilities for students and families with handicaps and special needs but little time and resources to pursue them outside of school facilities and regulations.

The evolution of performance measurements and rewards in higher education can be seen in terms of an historical evolution covering nearly four centuries of expanding expectations of colleges. A separate paper in this series provides a detailed accounting of this evolution of higher education and its relationship to the economy over four centuries, during which objectives and concepts of performance

²² Stigler, An Academic Episode...

changed dramatically. This spans an original role for colleges in religion and culture, later involvement in the development of agricultural technology, following later with support for science and industrial development, and more recently as a more universal element of workforce development. This perspective is important as it helps explain the changing nature of performance goals and expectations driving incentives for educational institutions in more recent times. To summarize that evolution, we can view the evolution of American higher education in terms of five historical periods. The central unifying elements of them are the changes in stakeholder incentives – how and why rewards for “better performance” have changed and with what effects.

- 1) **1636-1850.** Classical college education was a stable concept remaining for centuries in the economic shadows, an inconsequential corner of an overwhelmingly unskilled agricultural economy. For well over a century after Harvard’s emergence in 1636 as the first college in the American Colonies’ pre-industrial economy, there was virtually no connection between a college’s activities and the labor market for its graduates. Weak incentives prevailed.
- 2) **1860 – 1940.** By the time of the Civil War, industrialization was changing the economy and role of colleges. The Federal Land Grant College Act (1862) and its supplementation by the Hatch Act (1887) advanced colleges’ outreach goals from their origins as teachers of classical languages and literature to children of wealthy landowners, to conduits tasked with unifying teaching with research and applications to commercial enterprises – aided by Federal subsidies and relaxed restrictions on profits generated jointly with university and Federal support. Agricultural Experiment Stations at state universities, for example, assisted farmers and food processors to resolve problems of plant and insect diseases affecting agricultural business. In later years, colleges expanded their role in industrial development and training.
- 3) **1944 – 1965.** Following World War II, the first “G.I. Bill of Rights” in 1944 expanded college access to 12 million WWII military veterans. Higher education came to be seen as increasingly essential for the future of the masses, not just for a tiny wealthy elite. “Essential” at first for WWII veterans but later for broader groups, higher education expanded rapidly as a means for training a post-war labor force, though that occurred with greater costs. Costs were especially contentious through the 1950s and 1960s.
- 4) **1965 – 2020.** With expansion of information technology and white collar jobs, higher education came to be seen as increasingly relevant for workforce development. The tasks for higher education grew and expansion of higher education did much more than generate more and larger colleges. College tuition, public subsidy and loan programs also grew. With that, consumer and public demand grew for “better” *measurements* of college performance, raising needs for measurement while their systematic limitations have taken a toll. Stakeholders have become increasingly adept at gaming the reward system for private rather than social gain.
- 5) **2020’s and the Future.** New goals have been emerging in a dynamic economy that is both responding to the emergence of remote learning and the transformation of colleges from teachers of pre-existing ideas, into research institutions that are expanding patterns of global trade, improving workers health status and working-life expectancies, and raising labor-force-participation rates. So, too, have expectations, rewards, penalties, and controversies regarding colleges’ “social responsibilities” in such areas as equity, environment, and investment priorities. These issues raise the question of how to define, measure, and value “quality” of higher education, as those concepts expand.

As this evolution continues, there are increasingly important Issues of how to *measure* the degrees of realizations of a college's multiple goals, and how to strengthen the ties between the measurements and the financial rewards a school receives. are fundamental to establishing efficient incentives though they often bring unwanted side effects; success must not only be measured, but also converted into values, to establish appropriate rewards.

The Continuing Problem for Higher Education

Today's surging demand for higher education is far more utilitarian than that of colleges in the 17th, 18th, and 19th centuries, as there is a common focus on its contribution to earning power and economic growth. With this evolution has come greater government and citizen demand for evidence that public support for higher education passes a benefit-cost test -- justified by labor-market returns in excess of the education production costs.

Yet measuring and valuing performance in higher education are not problems that will soon disappear even as society's prowess with measurements improves. The race between developing "hard, replicable, numbers" to gauge performance, and stakeholders' incentives to game those measurements -- exaggerating favorable outcomes, minimizing unfavorable -- will continue.

When measures of performance are incomplete or biased -- under-rewarding some actions and over-rewarding others -- inefficient incentives result. Strong rewards for *measured* performance will encourage their provision; weak or even no rewards for *unmeasured* elements of performance will *discourage* their provision.

The growing number, variety, and interactive complexity of goals being pursued by colleges, students, government, and other stakeholders are calling for additional performance measures, each bringing new incentives. In the process, new opportunities emerged for the various stakeholders to game the reward mechanisms. The gaming is the result of each stakeholder's incentives to concentrate on performance measures that are easy to implement and so to garner larger rewards, while paying little attention to forms of performance that are costly and controversial to measure, value, and reward.

Gaming of a reward system is a result of the "multi-tasking" problem, the simultaneous pursuit of conflicting goals when greater success in achieving one goal reduces the success in achieving another goal. Optimization of the conflict requires measuring and valuing performance toward each goal, and the degree to which better achievement of one goal causes diminished achievement of another.²³ Earlier papers in this series²⁴ showed for other industries how rewarding and penalizing only *some* dimensions of performance, and essentially ignoring others, undermines overall efficiency. Thus, greater rewards for "better" performance, as measured, elevates incentives to game the rules. The same applies to higher education.

Higher Education "Performance" from the Student/Parents Perspective

Gaming of quality is especially troubling in higher education because it is hard to measure comprehensively, and to value, even when it is visible. The result is that stakeholders, governmental and private, have the incentive to tout measures of quality that advance the school's prominence -- such as

²³ Holmstrom and Milgrom, Multitask Principal Agent...

²⁴ Weisbrod, "Why Strong Performance...", Weisbrod, "The K-12 Education...", Weisbrod, "Financial Rewards for Health Care..."

the percentage of undergraduates who participate in a “study-abroad” program – while avoiding dissemination of measures such as student complaints about “inappropriate” remarks or behavior by other students or faculty. Both forms of collegiate self-reporting, one favorable, one not, illustrate the opportunities and biases inherent in reliance on the differential costs and benefits of any P4P reward system, whether it be a college, a hospital, a police department or any other service provider that is relied upon for performance information that will affect its rewards and penalties.

The previously discussed college rankings by *USNWR*, for example, are familiar to most graduating high school students and their parents, yet little information is disseminated publicly about the measurement process by which the rankings are determined, their justifications, their reliance on self-reporting by the colleges that have the incentives to pursue higher ranks, and efforts by *USNWR* to monitor and enforce their rules to avoid gaming. As a result of the widespread acceptance of the rankings as meaningful, worthy of being rewarded, the occasional evidence that, for example, a college president’s employment contract pays a bonus if the school’s rank improves, little is known publicly about the techniques being used to garner such rewards – and the gaming-induced distortions that result.

But there is no doubt about colleges’ *incentives* created by these rankings: they are easy to develop, low cost to implement, ways to reap the rewards for higher rankings and other elements of performance -- as they are *measured*.

Cost-effectiveness measures encourage colleges to reduce their *costs* in ways that do not affect *quality* in forms that are not easily recognized by students and parents as materially reducing “quality” -- larger class sizes, grade inflation, increased availability of course-credits for “life-experiences” that speed graduation, cut tuition costs, and substitute temporary and part-time, adjunct, faculty for more accomplished tenure-track faculty.

Output measurements and valuations are critical to the formulation of efficient incentives, but they also differ in their accuracy, degree of bias, and costs. As a result, they encourage stakeholders to use and tout the particular measures that bring them the greatest rewards.

So, the more *costly* it is for stakeholders in the college market to detect, measure, and value a school’s performance, including quality, the more perilous it is to use strong incentives to encourage *better performance*. *Pay-for-performance will encourage gaming, rewarding not what is truly important but what is easily demonstrated* -- as we saw this at work in the case of Volkswagen’s gaming of the U.S. testing of automobile emissions, as we saw in Paper #1, and with violation of the public school’s test-grading system in Atlanta, Georgia, as described in Paper #3.

In these cases, what was actually measured, tested, and rewarded was neither the *true*, unbiased, level of auto air-polluting emissions, nor of student performance on the standardized school tests; the *gamed* measurement procedures used by the auto firm and by the Atlanta Superintendent of Schools were reported and rewarded. But *true* performance was over-reported and, as a result, excessively rewarded.

For-Profit Colleges: Can they Game Governments, Students and Their Parents?

The allegation that for-profit (FP) colleges are especially suspect of fraud has in the past received attention in connection with recruitment of students with promises of well-paying jobs that will more than cover the student’s loan repayment debt for financing tuition costs. FP colleges have been linked to almost all loan fraud claims, according to an analysis of U.S. Department of Education data by The

Century Foundation in 2017. Students who attended FP colleges filed more than 98 percent of all requests for student loan-repayment forgiveness based on allegations of college fraud, The Century Foundation reported. The analysis covered the nearly 100,000 claims known as borrower defense received by the Department of Education over the previous two decades: “Of the more than 98,800 complaints received by the department as of mid-August [of 2017], 98.6 percent came from students at for-profit schools,” although those schools accounted for only about 10 percent of national enrollment and 18 percent of federal student debt in 2017. FP colleges can be responsible for student *borrowers’* claims of fraud and for reimbursing *taxpayers* who financed student loans.²⁵

Allegations of FP college fraud, and of the colleges incentivizing their student recruiters to aggressively pursue low-income, often academically weak, students to borrow from government student loan programs to pay tuition, have been common. Students who were unlikely to be able to repay their loans, and who could not escape the debt through bankruptcy, were allegedly being persuaded to borrow tuition money, which the college could generally retain. The process and its alleged abuses led, in 1992, to introduction of a Federal government restriction on FP colleges’ reliance on student loans from Federal sources. The new rule limited a FP’s reliance on student borrowing for tuition; a maximum of 85 percent of the school’s total revenue could come from Federal loans – hence the “85-15 percent rule.”

The “85/15” rule was arbitrary, and its restriction to FPs was strongly opposed by them, for it restricted their ability to generate added profit by recruiting more low-income students who financed tuition by borrowing from government loan programs. The FP colleges’ lobbying efforts succeeded, and the rule was soon relaxed to “90/10,” permitting the schools to recruit more low-income students who, with the help of the school, financed their tuition by borrowing from a Federal student loan program.²⁶ In effect, the Federal government, as a lender to students, was being gamed by FP colleges who were using low-income students as instruments for generating tuition revenue and profit.

What’s in a school’s name?

There could be a lot. What exactly is, or should be, a “college” or “university” eligible for government subsidization? An interesting case involved the establishment in 2017 of a new school, MissionU, which seemingly was capitalizing on its being a “U”. It was a traditional school in that it targeted students of typical college age (about 18-22), and they took courses. But it followed a dramatically non-traditional educational plan, offering only a single, one-year, “skills-focused,” program rather than a conventional four-year program. MissionU thereby cut traditional schools’ program and student costs, eliminating three of the conventional four years of time and tuition to get an education based on theory.”²⁷

MissionU was *non*-traditional in another sense. It had a very unconventional business model, which combined education with insurance, for the school generated revenue that was conditional on the student’s *future* income. Rather than paying a pre-determined tuition up-front, MissionU students were charged a fee contingent on their *subsequent* income; tuition was, in effect, a form of profit-sharing arrangement under which students were “expected to repay 15 percent of their income for three years upon completion of the [one-year] program -- but only if they earned a salary of at least \$50,000.” For its first student cohort of 25, the one-year program was in *data analytics and business intelligence*.

²⁵ Associated Press, for Profit Colleges....

²⁶ Clark, Wofford, and Ament, Stop For-Profit Colleges... ; Lee and Loone, Understanding the 90/10 Rule....

²⁷ Lederman, Self-Proclaimed Alternative...

Students had few if any course options. In year two, 2018, student applications were down sharply, and the school closed.²⁸

Should MissionU and other unconventional types of post-secondary programs that may be developed in the future be eligible for governmental financial support? If government pays a “school” for its performance “success,” how should that be measured for a school such as MissionU, and then converted into a monetary value to establish efficient incentives? Are total job market earnings over the next several years a significant component of a college’s performance? And what is the social value of a college having a higher “graduation rate”? Any tuition-payment arrangement that rewards a college for such performance, whether determined by government, students, colleges, or employers, can and will be gamed by stakeholders pursuing private rewards? Complex concepts.

Shackling a college’s revenue from tuition, donations, or other sources to observable measures of its performance, has become increasingly common. But the concept of Pay-for-Performance is old. It reflects a view, dating back to at least 1776 and Adam Smith’s insights in his book, *The Wealth of Nations*, into the foundation of a decentralized private market economy. What have come to be seen as inefficient “private market failures,” are forces driving wedges between private and social benefits, as well as between private and social costs; the result is that *individuals’* pursuit of self-interest does not lead to *social* efficiency.

Gaming of Government Tax Laws

U.S. tax laws encourage tax-deductible private donations to nonprofit colleges and universities (and other “charitable” nonprofits) under section 501(c) (3) of the Internal Revenue Code. In addition, nonprofits’ “endowments” and the profit from their investments are also largely untaxed, as are profits earned from the pursuit of tax-exempt activities. The various tax exemptions and deductions for private donors and for the recipient nonprofit schools – and consist very largely of private nonprofit and state-owned schools that are generally exempt from Federal taxation – so suppliers and consumers of higher education have the incentives to game the tax and regulatory systems, to shift financial burdens to other stakeholders --governmental or private.

One way a donor might game the tax system faced by nonprofit colleges is to under-report to the IRS the “fair market value” (FMV) of any goods or services the donor received in return for a “donation.” For the allowable tax-deductible for a donor is *not* the *entire* contribution, but only the excess of the donation above the FMV of what the donor received in return, apart from a “token” thank-you gift. The legal determination of what *token* means can be gamed. And it has been.

Not long ago a major Midwestern university gave a “thank-you” gift to a donor -- upgraded football season tickets, from near the “end zone” to near the 50-yard line, at its sold-out, 100,000 seat, stadium - in return for his “donation.” The FMV of the improved seating promptly became an issue for the IRS. True, as the donor argued, the prices printed on both the original and the upgraded tickets were the same. But according to the IRS that did not imply that their FMV was the same. The IRS determined that the FMV of the 50-yard line seats was greater, regardless of what the tickets showed as the price; the IRS was the tax law enforcement agency, and it won the dispute. A substantial tax bill to the donor resulted, sending a signal to future donors.²⁹

²⁸ Lederman, Self-Proclaimed Alternative...

²⁹ Colombo, The Marketing of Philanthropy...

In this case, two stakeholder groups – a nonprofit (NP) university and its associated tax-exempt foundation, and the private donor and football fan -- had a shared interest. It was to game a third stakeholder, the IRS as agent for the Federal tax system, by cutting Federal tax revenue, while reducing donors' tax liabilities and stimulating larger contributions to the university. The method was to find a way for (a) a donor to get a larger income tax reduction, (b) for the university to get a larger contribution, and (c) for the IRS to receive less tax revenue. The strategy for the university and its donors was to transfer net revenue from one group of stakeholders, taxpayers-in-general, shifting it to two other stakeholder groups. There was an incentive for both the college and the private contributor to game the tax system by understating the FMV of the better football seating the "donor" received.

The IRS confronted another hurdle in this struggle with other tax-system stakeholders, its limited and decreasing budget for enforcement. There was a strategic struggle between the IRS as a tax collector, and section 501(c)(3) of the tax code, which specifies tax-collection responsibilities for nonprofits. The IRS undertakes audits to uncover "errors" and "tax evasion" in the Form-990 informational returns required of most tax-exempt nonprofits. True, the "charitable" nonprofits are legally exempt from corporate profits taxation, but not from *all* profits, only those from activities that are "substantially related" to the nonprofits tax-exempt mission. What counts under the Internal Revenue Code is not how a nonprofit *uses* any profit it reaps, nor on whether the profit is used to advance the charitable mission, but on whether the profit-generating activity itself advances that mission. So a NP has two clear incentives -- to generate maximum profit and to minimize the amount of it that is subject to the Unrelated Business Income Tax (UBIT).

Nonprofits have yet another tax-related incentive, to *use* Generally Acceptable Accounting Practices (GAAP) to avoid, or at least to minimize, the taxation of profit from their UBIT activities. For example, consider the accounting allocations of *joint* costs, such as executive compensation and facility depreciation, which are incurred for both a nonprofit's taxable and tax-exempt activities. A tax-minimizing cost allocation for a nonprofit college would call for it to report as much as possible of the joint costs to the otherwise taxable UB output, not to the tax-exempt charitable activity, so as to minimize total taxation and maximize after-tax revenue.

The IRS resources available for enforcement of the tax laws, including audits of the IRS form-990 informational returns generally required from nonprofits, have been significantly cut over time. Between the years 2010 and 2017 the number of IRS auditors decreased a third, from about 14,000 in 2010 to 9,510 in 2017, and the number of audits fell 42 percent; the last time the IRS had fewer than 10,000 revenue agents was 1953, when the economy was less than one-seventh its current size.³⁰ The decreased enforcement budget for auditing the Form 990 informational returns has made it increasingly difficult for the IRS to determine, for example, when a nonprofit college or university is engaged in activities that are not *substantially related* to its tax-exempt mission, and so are subject to the Unrelated Business Income Tax (UBIT). Research has found evidence of such gaming, involving the allocations of a joint cost, depreciation, between nonprofits' taxable and untaxable activities, although the research did not specifically study colleges.³¹

The gaming of data on a college's "success" also highlights the differences between what *performance* means to stakeholders pursuing different private goals. Colleges, their tuition-paying students, governments as financiers and regulators, and private firms such as *US News and World*

³⁰ ProPublica, How the IRS Was Gutted...

³¹ Sinitsyn and Weisbrod, Behavior of Nonprofit Organizations...

Report, that design and implement information systems to rate and rank colleges and then sell the results; each has informational advantages and disadvantages over other stakeholders. Their measurements and school rankings can be manipulated, and they have been.

What “successful” gaming requires is development of a measurement methodology or an accounting, record-keeping, mechanism for tabulating and reporting performance. The challenge lies in the opportunities to develop measurements that (a) cannot easily be used in an unintended way that *increases* the measured social benefits or *decreases* the measured social costs, or that (b) “improves” or, at least exaggerates a school’s *quality* ranking – such as occurred in the cases of a school’s Alumni Giving Rate, discussed earlier in this paper. How often such socially inefficient gaming of measurements occurs, how long it persists, and the magnitudes of the effects, are seldom entirely clear, but the incentives to develop and use them, are.

There are many ways.

Governments as Stakeholders in Higher Education: Their Incentives to Game

One approach to gaming by a state government agency was adopted earlier this century by the University of Massachusetts at Amherst. Succumbing to political pressure to slow, if not reverse, the escalating in-state undergraduate tuition, the school found a “solution;” it succeeded without really succeeding -- by gaming the way “tuition” was *measured*.

The University’s success was quick and effective. Well, maybe.

Table 4.2 tells the story. It portrays a socially unwanted and economically inefficient way to control rising tuition. The mechanism was a form of price gaming; it was legal, and it worked for decades, but provided misleading information. But it also had a flaw that would prove fatal.

The University did succeed in suppressing the rate of increase of *tuition*; indeed, tuition fell. Yet the University *failed* to suppress the growth of what really counted -- *college-attendance costs increased, but the “tuition” component decreased*. The technique highlighted the importance of measurements and opportunities to game the measures. Altering the measure of, say, “tuition” can control the *form* of a cost while simultaneously failing to achieve the goal of cutting a student’s cost of attendance. The fundamental problem was the school’s incentive to manipulate measurements, in this case by re-defining key measures; whenever a reward hinges on the way a variable is measured and reported, a stakeholder’s incentive is to choose the option it finds most desirable, even if other stakeholders are misled in the process. I will explain.

In the 35 years between 1983-84 and 2018-19, remarkable changes occurred in the University of Massachusetts, Amherst’s, in-state student charges. The changes illustrate how a state government agency can and did game other stakeholders – in this case, students, their parents, and the general public.

The key goal was not to cut the official “tuition” as it was reported by the University, but to reduce the overall cost of a college education at Amherst. The two goals are not the same, and they are not equally costly to bring about.

Table 4.2, below, shows that in 1983-84, in-state full-time tuition at the University was \$1,129 per year, while mandatory fees were roughly another one-third, \$433. But ten years later, while tuition had increased in nominal, *not* inflation-adjusted, terms, had doubled, to \$2,220, mandatory *fees* had soared over seven-fold, to \$3,247.

The idea of reducing *tuition* but simultaneously increasing the *fees* students were required to pay, to at least hold the school's total revenue constant, continued for over 30 years, through 2014-15, when the evidence of purposeful gaming became public. *Tuition*, which had remained virtually constant for thirty years, increasing an average of only 2-3 percent annually, leaped **in** one year, from \$1,714 per year to \$13,790, while annual *fees* plummeted 97 percent, from \$11,544 to \$381, and tuition, which had been constant at \$1,714 for 16 years, leaped to \$13,790 in the single year from 2014-15 to 2015-16.

The gaming strategy was simple. "Tuition" was indeed being cut, or at least held constant in nominal terms, for more than 20 years from 1993-94 to 2014-15. Yes, but not really; tuition was essentially re-named. The strategic "solution?" Don't raise *tuition*, raise "*fees*."

It "worked." Tuition *was* indeed, suppressed, and the procedure was legal. But it was also a charade. The *totals* of tuition and *mandatory attendance fees* tell a very different story. Fees increased sharply, more than offsetting the reductions in tuition.

Table 4.2 Gaming: In-State Undergraduate Student Tuition and Required Fees, University of Massachusetts, Amherst, Selected Years

<u>Year</u>	<u>Tuition</u>	<u>Fees</u>	<u>Total</u>
1983-84	\$1,129	\$433	\$1,562

1988-89	1,404	1,114	2,518
1993-94	2,220	3,247	5,467
1998-99	1,904	3,325	5,229
2003-04	1,714	6,518	8,232
2008-09	1,714	8,518	10,232
2013-14	1,714	11,544	13,258
2014-15	1,714	11,544	13,258
2015-16	13,790	381	14,171
2016-17	14,590	381	14,971
2017-18	15,030	381	15,411
2018-19	15,406	481	15,887

Source: UMass, Amherst, Undergraduate Student Charges....

Summary "Take-Aways": Taxes Change Stakeholder Incentives, Encouraging Gaming

The college endowment excise tax is a process still unfolding. Yet its future contours, as identifiable forms of gaming, are already visible.³²

A. Timing: it matters when measurements are made. Key data that determine whether a college is subject to a tax, including the TCJA, and the magnitude of the tax liability, can depend on precisely *when* key data are gathered. It can matter a great deal, for example, precisely *when* a school counts the number of its students, and when and how the market value of its endowment is determined. These two statistics determine a college's endowment per student, and, so, whether the school is subject to the TCJA excise tax. A college can also discourage or encourage student registrations and withdrawals at

³² Lederman, Endowment Tax Picture...

particular times, and to influence the reported number of its students who can legally be counted to reduce the school's endowment per student.

B. Stakeholder coalitions: schools and students can game the bill-paying government. In the aftermath of WWII, a college and some of its students developed a creative way to game a third stakeholder, the Federal Government's Veterans Administration (VA). Under the 1944 G.I. Bill of Rights the VA paid college tuition for a WWII veteran who was honorably discharged if, and this was an important incentivizing force, the veteran stayed in a course until a formal "drop" date; if the course was dropped after that, the school could retain the full VA tuition payment.

At least one school found a way to game this VA performance-reward system. The school was paid for the combination of two dimensions of its performance: (1) getting more students to enroll, and (2) keeping the students in the program and its courses -- that is, deterring drop-outs -- until the VA rules permitted the school to retain the full course tuition.

This reward structure encouraged colleges to game the VA. Two stakeholders, schools and military veterans, shared an incentive -- to cooperate in a strategy against a third stakeholder, the Federal VA program, which paid the tuition. Each veteran who registered for the school's TV repair course, for example, received a "free" TV set as a "bonus" if the veteran remained in the course until the school could legally retain the full tuition.

The school, in effect, shared its profits with the students, at the expense of the Federal purse. The major goal of the G.I. Bill, though, was to expand education, not to provide free TV sets.

C. College Performance Measures: Gaming of Course Grades and Graduation Rates. These two common measures of performance in higher education illustrate the gaming hazards that strong rewards generate. College graduation rates can be increased, and lawfully, by, for example, inflating student grades (GPAs) and by making it easier and cheaper for students to get course credits for satisfying graduation requirements. These performance measures have significant attractions beyond their appeal to students and their parents; they are familiar, easy to trace over time, and easy to compare across schools and programs.

D. Cutting costs by shaving "quality." Educational quality is multi-dimensional and volatile over time, and the importance of each dimension differs among stakeholders and over time. Measurement problems abound for assessing and aggregating the diverse forms of benefits that are expected of higher education.

While measuring the varied and changing forms of social *benefits* from higher education, and their market values, are difficult and controversial, the forms of increased or decreased *costs* are typically easier to evaluate. So, if cost reductions are easy to observe, measure, and value, while their effects on "quality" are not, the end-result of joining "pay" with "performance" is predictable. The privately efficient way for a college to expand student access to higher education is to cut expenditures on *quality* in ways that, even if they are readily observed -- such as increasing class sizes and decreasing teachers experience, the effects on student "lifetime learning" are not; those forms and measures of reduced quality will be under-valued.

There is a conclusion and a lesson: to increase a college's measured performance, find ways to (a) cut costs, while (b) cutting quality in ways that are hard for students, their parents, and government stakeholders -- to recognize and value. It's all about gaming measurements.

E. Gaming: What should determine whether a school is eligible for subsidization? What exactly is a college or school? Nearly 50 years ago, in 1972, the issue of what a “school” is, surfaced in an intriguing way. Mitzi Shore and her husband, Sammy Shore, opened the *Comedy Store* on Los Angeles’ Sunset Strip. It “served as a talent pool for Johnny Carson, who often chose young comedians like Jay Leno and David Letterman to be guests on his ‘Tonight Show.’ ” Was the Comedy Store a “college” that qualified for Government tuition payments?

“We’re like a school,” Ms. Shore later told *The Los Angeles Times*, in 1994. In 1979 a group of The Comedy Store comedians had gone on strike for several weeks because they were demanding to be paid, although they were obtaining experience and training that increased their expected future earnings. After several weeks Ms. Shore, who had become the owner, agreed to pay them, although, according to Rick Newman, founder of a number of other comedy clubs, the Comedy Store was a “comedy university.”³³

Conventional universities had for centuries *not* paid their “students;” to the contrary, they collected tuition from them. After WWII and the widespread advent of commercial television, major, Division 1, universities, began paying star football and basketball players, legally or illicitly, whose notoriety brought colleges revenue from ticket sales, television contracts, and college endorsements of athletic clothing sellers such as Nike. Traditions and the NCAA (National Collegiate Athletic Association) rules on “amateurism” allowed limited compensation of athletes beyond tuition scholarships.

F. Educational “Quality:” Measuring and Rewarding It. Questions remain about determining a school’s eligibility for federal government tuition subsidies. What should and should not qualify as a “higher education” institution, and as a program or course “deserving” a subsidy? Should there be a *bright line* between a school’s educational and its non-educational activities? Should a school be permitted to *train* students in the use of occupational skills, collecting tuition in return, and at the same time to *sell* those students’ services, however limited, to clients who pay the school for the services? Should a college that teaches accounting, for example, qualify for government financial support even if it also sells the students’ tax-preparing services to retail clients?

The point is that a college can be, and often is, not only a seller of education, but also of non-education services such as income-tax preparation and provision of profitable football and basketball tickets including the annual “March-Madness” college basketball tournament? Expansion of the content of the higher education industry will continue to accelerate in the coming years, and with that, so will the debate over what its goals *should be*, and how they should be pursued? “Educating” young people is certainly high on the list, but that masks many questions, most of which involve the ambiguity of what “educating” should mean and how its measurement and valuation should change over time. Expanding *total* college attendance, expanding (or contracting) the *diversity* of student bodies, and what that should mean, increasing the diversity of faculty in racial, ethnic, wealth, and other forms, and maintaining if not increasing education “quality,” are other elements of the expanding goals for higher education.

At the core of competition among multiple goals are the connections between education quality and its *costs*. Costs are generally easy to measure; quality is not. Higher education *can* be provided to more students, and even without great cost stress, if society does not care about quality or how to gauge it. For example, hiring more low-wage, adjunct, faculty, increasing class sizes (student-faculty ratios),

³³ Slotnik. Mitzi Shor, Comedy Club...

expanding use of cost-cutting instructional technologies such as *online instruction*, and relaxing requirements for graduation, are ways to cut a college's costs per student -- but that does not imply that quality can be maintained with cost-cutting.

Change in college faculty composition in recent decades has been striking. The National Center for Education Statistics (NCES) has reported on the increase of part-time faculty. At all degree-granting, postsecondary institutions, between 1999 and 2016, *full-time* faculty increased 38 percent, but part-time faculty rose by nearly double that --74 percent -- from 1999 to 2011. Viewed another way, part-time faculty spurted from 22 percent of U.S. college faculty in 1970 to 50 percent by 2015, before leveling off to 44% as of 2022.³⁴

There are short-run (SR) and long-run (LR) effects of cost-cutting. The LR effects on educational quality and future job opportunities, for perhaps decades ahead, are largely unknown, but much depends on the ability of government and student stakeholders to predict, measure, and value changes in the forms of educational quality, and in the costs of preventing quality from being intentionally gamed by stakeholders whose self-interests encourage their manipulations of both performance measurements and their public reporting. Any form of quality that is either not measurable or measurable only at high cost, is a prime candidate for intentional misreporting of the performance measures that are *actually* rewarded, and not those that would *ideally* be rewarded. And what is not rewarded is *underprovided*.

Cost-cutting, Educational Quality, and the Growing Appeal of Adjunct Faculty

College administrators are aware of the incentives leading them to pursue increased revenue and avoid increased costs. Colleges also recognize the desirability of retaining faculty flexibility so that changes in students' fields of interest, labor market demands, or the availability of government grants in particular research areas, can allow the school to respond. In short, the *efficient* composition of faculty is not fixed over time, yet the *actual* composition of faculty can be rather rigid if the fraction of faculty having job tenure in a specific subject field remains "high" even while student demand has declined.

So, there are trade-offs among a college's goals. How much should it focus on (1) low *short-run costs*, and particularly on faculty salaries and teaching responsibilities, on (2) high *long-run quality* -- assuming that this entails commitments to tenure-track faculty who have achieved high standards of professional achievement and are rewarded by higher pay, and on (3) *long-run* commitments to academic leadership in advancing new directions for research and teaching.

There is a potential "solution: costs of some education might be cut in a way that has little or no effect on *revenues*. But how hard is that? Part is easy, part is not; therein lies the problem of linking pay with measured performance.

An example involving the choice between high-cost and low-cost faculty highlights the dilemma. When less-costly adjunct faculty can be substituted for higher-cost, tenure-track, faculty, and the shorter-term employment contracts facilitate a school's adaptability to changing programmatic demands over time, does *quality* suffer in the process?

Measurability is the key. The facts that Adjunct Faculty are typically part-time and temporary, that they teach primarily undergraduates and are not judged on their performance through academic *research* and publications, and that they have greater teaching "loads," larger classes, and lower

³⁴ Way, With Colleges Shifting... ; NCES, Characteristics of Post-Secondary Faculty...

salaries, are double-edged swords -- all are easily observed and measured, but other elements of college performance are not. Some of these “immeasurables” may be captured by various “quality” variables, though there is uncertainty about how to measure them, other than by individuals’ opinions, which encounter other hurdles.

A college is rarely judged solely by its prices such as tuition and fees, net of student financial aid, or by other *elements of performance* that can be captured by market prices. These immeasurables, though, are at the root of this set of papers, on *The Perils of Pay-for-Performance*. Being hard or even impossible to measure or value, when some elements of performance are rewarded little, if at all, delivers a powerful blow to efficient performance decisions.

Faculty labor costs. Some 50-60 percent of total operating costs in higher education are for faculty, and so, serious cost-cutting efforts are bound to focus on them. If students and parents do not perceive Adjunct or part-time teachers as compromising education quality, and insofar as government programs to expand student access to college do not pay colleges less if they employ more of them, colleges have the incentive to game the pricing system, employing more of the less-costly faculty. Substituting lower-priced faculty is especially appealing if students, their parents, and government agencies do not interpret the lower-cost faculty as providing lower quality education.

Looking to the future, we can also easily see part-time and limited-term faculty appointments as increasingly appealing to a higher education industry facing growing fiscal stress and uncertainty of demand for traditional forms of schooling. Competitive pressures to experiment with uncertain demand for non-traditional programs and methods often do not have staying-power, causing faculty contract terminations and replacements with part-time, limited-duration appointments. It is helpful, though, especially to the hundreds of colleges with no significant endowment wealth, to develop and fall back upon temporary “rainy-day” funds to support innovative cost-cutting devices.

Another way to cope with a college’s fiscal stress is to Increase course credits toward graduation by giving students credits for past “life experience,” with each college exercising discretion on specific requirements, and with the result that fewer formal courses, less calendar time and less revenue from student tuition are required for graduation. Making courses, exams, and graduations easier can also speed the graduation process, thereby strengthening a college’s competitive position, although perhaps by degrading quality.³⁵

More new forms of gaming and cost-cutting are ahead. For there is no consensus on how to measure the many dimensions of “quality” of a college education, or on what activities other than traditional classroom-based or, perhaps, online courses -- a college should be permitted to award graduation credits for, or what a college should be permitted to use as a basis for awarding “life experience” credits? In the competitive struggle for students and revenue, simple, even simplistic, performance measures such as a school’s 4-year or 6-year graduation rate, can retain their attraction even while an elusive educational *quality* fades away as a serious consideration.

Sports and Education Performance Measures

In the early 2000s the National Collegiate Athletic Association (NCAA) responded to charges of poor performance by its member universities -- especially the high “dropout rates” or low “graduation rates”

³⁵ Wolverton, *The New Cheating Economy...*

of student-athletes in football and men's basketball. It developed a "solution" that was simple and effective, but it was gamed, and accomplished little.

The NCAA simply changed the way teams' *graduation rates* were measured. The newly invented, renamed, and re-measured "Graduation Success Rate" (GSR) was an immediate "success;" change the way "performance" was measured, and watch it leap. Easy.

The key was to count only those student-athletes who (a) enroll as freshmen, (b) receive athletics-based financial aid from that school, and then (c) determine the fraction of them that graduate from *that* institution within six years of initial enrollment.³⁶ (d) Do not count transfer students to or from the school. Simple.

The "improvement" in student-athletes' measured performance was dramatic. At the University of Florida, for example, the Federal government's new measure of the 6-year graduation rate for men's basketball players who enrolled anytime between 1997-1998 and 2000-2001, was 67 percent; but by the NCAA's new GSR measure it was 100 percent!

For football team players the difference was even more striking-- the 35 percent graduation rate, according to the Federal government's calculation, more than doubled to 72 percent, by the NCAA calculation. And for the University's baseball players, the Federal Government measure, showing that 26 percent graduated within six years, leaped to 71 percent by the NCAA's GSR metric.³⁷

That was one way to improve a school's *performance* – academic or other; change the way it is measured, a technique that also applied to "tuition" and "fees" at the U Mass, Amherst, as table 4.2 showed. After all, there is no single, "correct," way to measure a college's "true" *tuition* or *fees*, or its basketball team's *graduation rate*. So, there is fertile ground for gaming the measurement process.

True, there are other paths to "improved performance" beside measurement manipulations. But they are generally more costly, especially when the costs of policing, monitoring, and litigating are included, not to mention the side-effects of successful gaming. College athletic coaches *could* be incentivized away from maximizing the number of games won, and rewarded for encouraging players to devote more effort to academics and speedy graduation. They could, but making that work without serious consequences would be no small feat.

If "better" performance is to be encouraged through incentives, there is no way to avoid such basic questions as: What is meant by the "higher education" to be rewarded? How should its success be measured and valued? By which "providers" may it be produced to qualify for government performance-based rewards? What "educational outputs" and sub-par "quality" should and should not be eligible for what performance-based rewards? What are the consequences of paying differential rewards for varied forms of performance that differ in their measurability -- for example, whether a student correctly solves a math problem is easier to determine than is the "quality" of a student's artwork.

Rewarding performance is essentially a problem of writing a "complete" contract -- one that specifies (a) all rewards and penalties, (b) to each stakeholder, and (c) for every combination of

³⁶ Weisbrod, Ballou, and Asch, *Mission and Money...*

³⁷ Weisbrod, Ballou, and Asch, *Mission and Money...*

performance dimensions. For rewards and penalties to be efficient they must be aligned with the unbiased measurements of a school's performance. No small task.

Loan Repayment and the Moral Hazard Problem

Student loans become a seemingly low-cost way for government to support "Higher Education for All" without "Breaking the Bank." Because the series of subsidized Federal loans to students since WWII required repayment with interest, although at below-market rates, and because borrowers' repayment obligations were not necessarily extinguishable by low-income loans or even bankruptcy, cuts to government grants became more politically attractive.

Yet the loan burdens, repayment defaults, threatening collection tactics, and other "abusive lending practices by companies that administer federal loan programs" ³⁸ were more than the downsides of dependence on loans to expand higher education without breaking the bank. The door to the next phase in higher-education finance remains open.

Student Subsidies and Incentive Thresholds

Gaming a reward system is more than an abstract possibility; it should be expected, since it is a natural outcome of the pursuit by stakeholders of their own self-interests. Moreover, the greater the incentive is for gaming -- whether legally or not -- the more it will occur, and the more it occurs the more the public policy goal of linking rewards to *true* performance will be undermined. Stronger rewards are the seeds of stakeholder gaming.

A college that is compensated, directly or indirectly, by government payments for students faces two types of financial incentives to game the reward system. First, it has a *marketing incentive* -- to attract more students, even very weak students, as long as the school receives compensation sufficient to more than cover its costs: It also has a *retention incentive* -- to keep students in a course long enough for the school to keep the tuition payment, whoever pays it, beyond that retention date; with the full course tuition in the college's hands, if a student drops out there would be no financial loss to the college, even if its reputation suffered from, say, its "low" graduation rate -- more about that later.

So, there are strong incentives for a school to recruit and retain profitable students, but once a school has done what is necessary to retain the full tuition, a school's financial incentive to keep students in the course drops precipitously, perhaps to zero.

The day before that threshold drop date, the financial reward for a school's better performance -- retaining the student longer -- is the entire tuition for the course; on that date the school's reward shoots up to the full tuition; after that, the reward for retaining the student longer again drops to zero.

With such a threshold-based reward, a school's financial incentives are clear, but economically inefficient. They imply that (a) until the reward threshold is reached, an increase in a school's measured performance -- retention of students -- is not rewarded at all, through any added tuition income -- but (b) at the point the threshold is reached, even if it is considerably prior to the end of the term, the financial reward to the school becomes the entire tuition for the course, and (c) if a student remains in the course beyond the tuition-retention threshold, the school receives no additional reward, seemingly assuming that a student learns nothing more by remaining in the class, and so the school receives no additional compensation for increased student retention. Just below that tuition-retention threshold the

³⁸ Thrush, Trump Administration Battles...

reward system is powerful -- a tiny increase in measured performance brings a major increase in the reward -- while at other potential dropout dates, either before or after the threshold date, the reward system is exceedingly weak, providing no incentive for schools and faculty to work to enhance student achievement by encouraging them to continue in the program longer, whether or not the reward threshold is crossed.

The battle between expansion of government expenditures to increase student access to higher education, and government budgetary cost containment was on. So was stakeholders' struggle to become more inventive in their responses -- to game the reward system.

Student loans for higher education were being superseded by alternatives including non-repayable educational grants, as well as an expanding conception of what forms of education should be eligible for government grants. As competition among colleges has grown in the search for revenue, new courses and programs have been introduced, highlighting such questions as what the limits of higher education awards should be, what should count for a student's credits toward graduation, and how many such credits should be required for graduation and other forms of performance certifications ?

The Social Heisenberg Principle in Higher Education: Measuring and Gaming Performance Distorts Incentives

For anyone convinced that strong rewards for better performance are the keys to economic efficiency, in any part of the economy, the *Social Heisenberg Principle* is more than sobering. It is a troubling reminder that Pay-for-Performance -- aligning stakeholder rewards with their *measured* performance -- in higher education or elsewhere -- is essentially doomed from the onset, because the act of measurement distorts the performance being measured to align with incentive-inducing rewards.

In Paper 1 of this series, I termed this the Social Heisenberg Principle, derived from Werner Heisenberg's finding in the world of physics -- that it is impossible, not just costly, to measure, simultaneously, both the *location* of a subatomic particle and its *rate of motion*, because the process of measuring either one affects the other. In the world of higher education, the analog is that the process of rewarding measured performance affects the system that rewards it -- so the more-powerful the reward for performance is, the stronger is the incentive for a stakeholder to game the measurement process by overstating or understating the true performance.

Higher education today is an industry transitioning to stronger, more powerful, incentives. More distortions lie ahead as a college's "performance" shifts from one based on students' or administrators' impressions and judgments, to more "scientific" measures that are replicable and comparable across schools and over time, and that are measured in unbiased ways. The social goals, after all, are to achieve meaningful results, not to provide mismeasured and misleading numbers; stronger rewards are not the solution to the need for greater efficiency and expenditure control; they are often the problem.

Massive change is now underway in the incentivizing of better performance in higher education. The process is spotlighting the distortions that the Social Heisenberg Principle is bringing to the industry, and what lies ahead -- the content of higher education, the comprehensiveness of access to it, and its quality. They are not predetermined today, nor were they when the first colonial college opened in the Boston area nearly four centuries ago. Much of the spirit of the challenges is captured in a recent article in the Boston Globe, published in December of 2024 : "Too much student debt? Or too many students in college? Conservatives take aim at higher education."

Then, as now and, doubtless, in the foreseeable future, incentives as determined by the Pay-for-Performance (P4P) rules, mattered. They were very weak in Colonial days, and so they mattered little. Today they are increasingly strong, so they matter far more to all stakeholders, and they are affecting stakeholder incentives.

1. *For students and their families*, incentives operate through the costs of attending a particular college or university, including tuition, fees, and housing costs, and the benefits of attending, particularly the distribution of increased future earnings resulting from the education.

2. *For colleges and universities* their incentives and choices encompass decisions to enter or exit, what tuition and fees to charge, and what programs and faculty to provide. Schools can also choose their efforts to generate revenue from various sources such as tuition, fees, private donations, business corporations, and profits from business activity that is not “substantially related” to their tax-exempt mission -- the “Unrelated Business Income” (UBI) that is subject to ordinary corporate profits taxation, but may nonetheless generate profit. Nonprofit organizations, including hospitals, and other “charities” -- specifically those exempt from corporate taxation under section 501(c)(3) of the Internal Revenue Code -- also benefit from other Federal and state tax laws that exempt them from taxation of capital gains and income from endowments.

3. *Governments* influence colleges and student stakeholders in still other ways, providing access to subsidized student loans, to research and other grants to colleges, and to tax deductions for private donors to nonprofit schools. But there is a downside; by offering these incentives, governments unintentionally alter schools’ tuition pricing practices, endowment accumulation and investment policies, and incentives for other stakeholders to game the higher education revenue and expenditure system; some examples:

Students may, and do, default on their loan repayments, although they often cannot extinguish those debts even through bankruptcy. Students can also choose to spend more freely on higher-tuition schools and higher-quality living accommodations, knowing that they can borrow at a submarket interest rate.

Colleges can raise their tuitions and fees, knowing that subsidized loans are available to their students (This has been referred to as the Bennett Hypothesis, named for the former U.S. Secretary of Education, who hypothesized this counterproductive effect of government subsidies on college tuition.)

In short, government subsidies, while having the ostensible goals of cutting college tuition costs and thereby expanding access, should be expected to be systematically undermined as schools respond to new incentives to raise tuition, fees, and other user prices.

The Future of Performance Measurement and Incentives for Higher Education in the 21st Century

Today, higher education is rarely considered a luxury for the elite; the issue is how to make it financially accessible to “all.” “Earning a post-secondary degree or credential is no longer just a pathway to opportunity for a talented few,” the White House Web site asserted in 2015; “Rather, it is a prerequisite for the growing jobs of the new economy.”³⁹

³⁹ Cassidy, *College Calculus: Where’s the Real Value...*

Perhaps. But whether an industrialized economy does or does not require an increasingly college-educated labor supply, begs a number of questions. All involve ambiguities and with them, opportunities for each stakeholder group to game the available rewards by shifting the burdens to others or even to themselves in ways that are not recognized -- reduced "quality" in forms that are hard-to-measure or to evaluate; more about these later.

What exactly should be meant by "access" to a "college education?" How should it be financed, and of what "quality"? With potential students wanting to expand access, program choices, and quality, with colleges and instructors wanting greater access to governmental financial support for students, and with governments wanting to contain expenditures, the tensions among stakeholders are real.

Are students and their parents the direct consumers of college education? The central dimensions of their evaluations of a college's performance are the students' and parents' private costs and expected benefits: on the cost side there are tuition and fees, textbooks, and room-and-board expenses; and on the benefit side there are the "quality" of a specific school's educational "performance," the likelihood of a prospective student's graduating or, at least, completing a certification program, and post-graduation employment and earnings experience, as well as the program choices from which students may select.

From the student/parent – consumers – stakeholder perspective, lower costs, higher quality, and greater choice are preferred. But not so for other stakeholders such as producers (colleges) and governmental payers. For them the view is different; since there is no "free lunch," someone must pay.

Producer/College Stakeholders. For a school that provides higher education, satisfying consumers' goals of low prices and high-quality signals fiscal trouble. Lower prices (tuition, fees, etc.) paid by students mean less revenue for the school; higher quality implies more expenditures. This conflict between the interests of consumers and producers sets the stage for races among colleges to find both new sources of revenue and ways to cut expenditures while maintaining quality.

Taxpayer Stakeholders. There is another stakeholder group – taxpayers and the political leaders who represent them in the process of developing and operating the nation's higher education policies. The problem is that the three races – to cut tuition, to grow college revenues, and to cut tax burdens on the overall citizenry, conflict.

Multi-Dimensional Performance Measurement, Rewards, and Gaming Each stakeholder has the incentive to shift cost burdens to others, and to retain the benefits. Yet there are restrictions on their successes. Colleges typically have informational advantages over their student-consumers and governmental payers. True, a college can cut expenditures on faculty and facilities, permitting lower tuition, but if students and their parents recognize a deterioration of quality there will be consequences for the school. If at least some consumers or payers are not expected to recognize the decline in educational quality, colleges will take advantage of their informational superiority to shave educational quality without suffering adverse consequences.⁴⁰

⁴⁰ This is essentially the informational-asymmetry problem highlighted by George Akerlof, 1970, and illustrated by the used-car market in which, at least in the context of the state of automobile testing technology at the time, put the seller of a used car in a position of informational superiority over a prospective buyer. The analogous situation in another market, for human blood transfusions, was identified by Richard Titmuss in 1971.

The three central stakeholders in higher education – producers, consumers, and government – are in a strategic struggle to advance the social goal of expanding access to quality education while shifting as much as possible of the cost to other stakeholders. Their successes hinge substantially on their informational advantages and deficits compared with the others. These differentials provide opportunities to game the reward system. Information is power.

The future of performance measurement and rewards, while by no means entirely clear, is surprisingly evident because many of its elements have already surfaced. The changes flow from a small number of identifiable forces:

- (1) how to provide post-secondary education and training to essentially all who want it, which entails (a) defining the target population -- for example, only “young” people, or also mid-career and other adults? And (a) defining the contents of the education and training programs, and changing them as science, technology, and labor force opportunities evolve over time:
- (2) how to define, measure, and maintain the quality of programs in higher education.
- (3) how to accomplish (a) and (b) at “reasonable” cost; and
- (4) how to minimize the consequences of the inefficient gaming that will occur increasingly as incentives change.

Looking to the future, it appears clear that simplistic, single dimension performance measures with direct financial incentives, will encourage more gaming, but the undesirable outcomes can be anticipated and minimized. This will require a more nuanced combinations of multi-dimensional performance metrics, use of weaker incentives, and more attention to prevention of gaming.

Thus, the P4P incentive process is all about performance measurements and valuing their costs and benefits –fully, accurately, and without systematic bias. Stronger rewards are not equally effective in all industries, nor over time, as the states of knowledge, measurement techniques, and regulatory enforcement costs change. The more distortionary the effects of gaming are, however, and the more costly it is to detect and deter them, the more potent is the case for reducing, if not eliminating, use of strong rewards; *weaker* or *no* rewards may be more suitable.

This leads to a powerful conclusion about how to apply performance rewards in the higher education context. In an efficient economy, public policy will and should utilize a broad array of strong and weak rewards to encourage “better” performance. Strong P4P rewards, offering “larger” rewards for *measured* accomplishments, are most efficient when *true* performance is easily and accurately measured, and when it is costly to game the reward system without being detected and severely penalized. Weak rewards, by contrast, are most efficient when performance is complex, multi-dimensional, costly to measure and value, but easy to game and difficult to detect and prevent.

References

- Akerlof, George. "The Market for 'Lemons': Quality Uncertainty and the Market." *The Quarterly Journal of Economics*, Vol. 84, No. 3 (Aug.) 1970, pp. 488-500. <http://www.jstor.org/stable/1879431>
- Associated Press. "For-Profit Colleges Linked to Almost All Loan Fraud Claims," November 9, 2017. <https://www.cbsnews.com/news/study-most-student-loan-fraud-claims-involve-for-profits>.
- Burns, Hilary. *Boston Globe*. "Too much student debt? Or too many students in college? Conservatives take aim at higher education," December 18, 2024.
- Cassidy, John. "College Calculus: What's the real value of higher education?" *The New Yorker*, September 7, 2015. <http://www.newyorker.com/magazine/2015/09/07/college-calculus>.
- Clark, Wesley K., Carrie Wofford, and Aaron Ament. "Stop for-profit colleges from preying on veterans." *Chicago Tribune*, March 4, 2017: A17. <https://www.chicagotribune.com/2020/03/03/commentary-stop-for-profit-colleges-from-preying-on-veterans/>
- Colombo, John D. "The Marketing of Philanthropy and the Charitable Contribution Deduction: Integrating Theories for the Deduction and Tax Exemption." *Wake Forest Law Review*, 36(3), 2001, 657-700. https://www.researchgate.net/publication/228231542_The_Marketing_of_Philanthropy_and_the_Charitable_Contributions_Deduction
- Gifford, Storm "University of Oklahoma fudged alumni endowment rates for 20 years: magazine." *New York Daily News*. May 23, 2019.
- Golden, Daniel, "Math Lessons: To Boost Donor Numbers, Colleges Adopt New Tricks – Sinking Alumni Stats, Zeal for Rankings Supur Rate Inflation," *Wall Street Journal*, March 2, 2007, A1.
- Hamid, Rahem and Elias Schisgall. Harvard Report Shows 79% A-Range Grades Awarded in 2020-21, Sparking Faculty Discussion, *Harvard Crimson*, October 5, , 2023. <https://www.thecrimson.com/article/2023/10/5/faculty-debate-grade-inflation-compression/>
- Holmstrom, Bengt and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, Vol. 7, 1991. pp. 24-52.
- Kiel,Paul an Jesse Eisenger. How the IRS Was Gutted, *ProPublica*,December 2018. <https://www.propublica.org/article/how-the-irs-was-gutted>
- Lederman, Doug. "Self-Proclaimed Alternative to College Closes After a Year." *Inside Higher Education*, May 23, 2018. <https://www.insidehighered.com/digital-learning/article/2018/05/23/missionu-self-styled-alternative-higher-education-closes-after>
- Lederman, Doug "Endowment Tax Picture Becomes Modestly Clearer." *Inside Higher Education*, June 20, 2019a.. <https://www.insidehighered.com/news/2019/07/01/us-provides-some-clarity-about-tax-endowments>
- Lee, Vivien and Adam Looney. Understanding the 90/10 Rule: How reliant are public, private, and for-profit institutions on federal aid?, *The Brookings Institution, Center on Regulation and Markets*, 2019. https://www.brookings.edu/wp-content/uploads/2019/01/ES_20190116_Looney-90-10.pdf
- National Research Council. Improving Measurement of Productivity in Higher Education. The National Academies Press, 2012. <https://nap.nationalacademies.org/download/13417>
- NCES: National Center for Education Statistics, Degrees Conferred by Postsecondary Institutions, Table 318.10, 2023. https://nces.ed.gov/programs/digest/d23/tables/dt23_318.10.asp?current=yes
- NCES: National Center for Education Statistics,. Characteristics of Postsecondary Faculty, 2024. <https://nces.ed.gov/programs/coe/indicator/csc>

- Patel, Vimal. "How One University Measured Faculty Productivity," *The Chronicle of Higher Education*, February 29, 2016
- Rojstaczer, Stuart and Christopher Healy. "Where A is Ordinary: The Evolution of American and University Grading, 2004-2009." *Teachers College Record*, no. 7(2012), 2012: 1-23. <http://www.tcrecord.org/PrintContent.asp?ContentID=16473> .
- Sinitsyn, Maxim and Burton A. Weisbrod. "Behavior of Nonprofit Organizations in For-Profit Markets: The Curious Case of Unprofitable Revenue-Raising Activities." *Journal of Institutional and Theoretical Economics*, 164(4), December 2008: pp. 727-750.
- Slotnik, Daniel E. "Mitzi Shore, 87, Comedy Club Owner Who Fostered Rising Stars in Her 'University.'" *New York Times*, April 13, 2018, Obituaries.
- Stigler, George J. "An Academic Episode," Chapter 1 in Stigler, *The Intellectual and the Market Place, and Other Essays*. The Free Press of Glencoe, A Division of The Macmillan Company: NY, 1963: 1-7.
- Thrush, Glenn. "Trump Administration Battles States' Efforts to Police Student Loans." *New York Times*, September 7, 2018: A18.
- Titmuss, Richard M. *The Gift Relationship: from Human Blood to Social Policy*. Vintage Books, a Division of Random House, NY, 1971.
- UMass, Amherst. *Undergraduate Student Charges*, Office of Institutional Research, Budget Office, 2017; *Undergraduate Student Costs*, 2019.
- U.S. Census Bureau Current Population Survey, Educational Attainment in the United States: 2021 table package, Annual Survey and Economic Supplement, February 24, 2022.
- Way, Dan E. "With Colleges Shifting to Adjuncts, Teaching Quality May Suffer," June 22, 2018. James G. Martin Center for Academic Renewal, 2018. <http://www.jamesgmartin.center/2018/06/with-colleges-shifting-to-adjuncts-teaching-quality-may-suffer/>
- Weisbrod, Burton, Jeffrey Ballou, and Evelyn Asch. *Mission and Money: Understanding the University*. New York: Cambridge University Press, 2008.
- Weisbrod, Burton, "Why Strong Performance Rewards in Government and Nonprofit Programs Don't Work -- A Social Heisenberg Principle," paper 1 in the series: *The Perils of Pay for Performance of Public Service Industries*, Northwestern University Institute for Policy Research, Working Paper 25-01, January 2025. <https://www.ipr.northwestern.edu/documents/working-papers/2025/wp-25-1.pdf>
- Weisbrod, Burton, "The K-12 Education Revolution – The Perils of Not Just Measuring Performance by Standardized Tests, but Financially Rewarding It," paper 2 in the series: *The Perils of Pay for Performance of Public Service Industries*, Northwestern University Institute for Policy Research, Working Paper 25-05, February 2025. <https://www.ipr.northwestern.edu/documents/working-papers/2025/wp-25-5.pdf>
- Weisbrod, Burton, "Financial Rewards for Health Care Program Outcomes; How They Can and Do Generate Undesired Results," paper 3 in the series, *The Perils of Pay for Performance of Public Service Industries*, Northwestern University Institute for Policy Research, Working Paper 25-09, March 2025. <https://www.ipr.northwestern.edu/documents/working-papers/2025/wp-25-9.pdf>
- Wolverton, Brad. "The New Cheating Economy," *The Chronicle of Higher Education*, August 28, 2016. <https://www.chronicle.com/article/the-new-cheating-economy/>