Towson University Department of Economics **Working Paper Series**



Working Paper No. 2010-14

The Allocation of Merit Pay in Academia

By Finn Christensen, James Manley, and Louise Laurence

July, 2010

© 2010 by Authors. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Allocation of Merit Pay in Academia

Finn Christensen, James Manley, and Louise Laurence*

This paper investigates whether the widespread awarding of faculty merit pay at a large public university accurately reflects productivity. We show that pairwise voting on a quality standard by a committee can in theory be consistent with observed allocation patterns. However, the data indicate only nominal adherence to a quality standard. Departments with more severe compression issues are more likely to award merit pay as a countermeasure and some departments appear to be motivated by nonpecuniary incentives. Much of the variance in merit pay allocation remains unexplained. These results suggest reform is needed to improve transparency in the merit system.

JEL Classification: D7, I20, J33, M52

Finn Christensen (Corresponding Author) Towson University Phone: 410-704-2675 Fax: 410-704-3424 fchristensen@towson.edu http://pages.towson.edu/fchriste

James Manley Towson University Phone: 410-704-2146 Fax: 410-704-3424 jmanley@towson.edu http://pages.towson.edu/jmanley

Louise Laurence Towson University Phone: 410-704-2118 Fax: 410-704-3664 Ilaurence@towson.edu

^{*} We thank Tim Sullivan and James Clements for providing us with needed data. Errors are our own.

1. Introduction

Merit pay in organizations is designed to induce high effort by rewarding productivity. In educational institutions the goal is the same but the implementation is highly questionable, as on many university campuses merit pay is awarded too broadly to be considered much of an incentive. For example, surveys in Florida in 1992 found that two-thirds of faculty members received some form of merit pay (Anderson 1992). In 2007 part-time lecturers sued the University of Washington and obtained a settlement for that school's failure to distribute annual salary increases that went by the name "merit pay" (Gravois 2007). At the large public university which provided anonymous data for this study, the typical department awarded the highest level of merit ("merit plus") to twothirds of its members (Table 1). Why is merit pay awarded so generously in academia? Are recipients deserving?

Both of these questions have important implications for the design and effectiveness of merit pay on college campuses. If we assume that merit pay accurately reflects relative productivity, the fact that merit pay is awarded generously may imply that merit standards are too low, and an increase in these standards may elicit more effort from faculty on the margin. If merit pay reflects productivity poorly, then this suggests a lack of clarity for faculty on the relationship between output and rewards, or that faculty may be incentivized to spend time and resources on unproductive activities to attain greater merit pay. In either case, better alignment of merit and productive output would increase the effectiveness of any merit system from the principal's (i.e., the university's) perspective.

Award Year	# faculty in sample	Avg % Merit Plus – All faculty	Avg % Merit Plus - Tenured Faculty	Avg % Merit Plus - Untenured Professors
2007	492	67%	77%	64%
2008	519	68%	79%	59%
2009	563	68%	80%	58%

Table 1: Merit Plus Allocation by Tenure Status and Year

A unique aspect of academia that distinguishes it from most other organizations is that professors decide on merit pay for their immediate colleagues. Since professors are the experts in their fields, no one else on campus is fully able to evaluate their work. At the same time, asking them to assess their own productivity when they are aware that their evaluation will be linked to their remuneration clearly represents a conflict of interest. Such conflicts are intolerable in most sectors; even CEOs are often required to at least make a show of obtaining outside evaluation, though in fact they often exert *de facto* control of their own salaries (Elhagrasey, Harrison, and Buchholz, 1999).

That said, academic merit pay is not a free-for-all. Departments must justify their merit decisions since the dean or the provost has the final word. To this end, merit is typically tied to annual reports, which detail a faculty member's scholarly activity, teaching performance, and service for the past academic year. A person's merit pay is ostensibly some non-decreasing function of these productivity measures.

At the university we investigate, each department is assigned a pool of merit money to allocate among its members in the form of merit pay. In most departments, merit money is allocated by a committee (the *merit committee*) composed of the department's tenured faculty members, who must decide whether each department member has earned either "base merit" or "merit plus" designation. "Base merit" designation carries with it one share of the department's total award money, while "merit plus" represents two shares.¹ According to written policy, these decisions should be made solely on the basis of each faculty member's research, teaching, service, and in some departments "collegiality." Since the pool of merit money is fixed, awarding one person merit plus means that the remaining department members receive less. Thus, provided that merit decisions must be justified to a third party, each professor desires a standard of merit just

¹ A third designation, "no merit," also exists but it is seen as punitive and is extremely rare. Less than 2% of faculty in a given year were assigned "no merit" status.

loose enough that he is deemed meritorious. Any looser standard would erode the monetary and hedonic value of such a designation.

Another crucial aspect of the university we investigate is that there are separate policies for merit and retention. In principle, if an individual receives outside offer, and the university retains the individual by offering a retention raise, this may make it less likely an individual receives a merit raise. On the other hand, it is plausible that merit raises could be used to deter outside offers. Similar reasoning applies to how compression issues may influence merit decisions. (Compression refers to situations in which the salaries of senior faculty are low relative to the salaries offered to new hires). However, nowhere in the written policy on merit allocation is there any suggestion that merit decisions should depend on outside offers, retention raises, or compression. We therefore ignore these issues in our theoretical analysis.

Within the context of these institutional details, we show in a theoretical model that that the fact that a large share of the faculty are awarded merit plus can be consistent with sincere voting on an objective merit standard. With sincere pairwise majority voting, any standard may be chosen depending on the order in which the merit committee considers them, as in McKelvey (1976). With the admittedly strong, yet plausible and intuitive assumption that the committee proceeds through the available standards from strictest to loosest, we show that the committee will award merit plus to at least half its members.

We turn to the data to help us determine whether there exists a clearly defined merit standard, and if so, whether voting on a merit standard can explain why merit is so generously awarded. We are fortunate to have access to two sets of anonymous data. The first, a set of university-wide data, contains just information on faculty rank, department, and merit status in each of three years. A second set of data from a single college at the university contains both merit decisions and productivity measures.

5

At a minimum, use of a quality standard implies that the probability of getting merit plus is increasing in observable output, a prediction we verify with the college-level data. However, we observe that only a small amount of the variation in merit plus awards can be explained by variation in observable output. Moreover, the way in which observable output influences merit decisions differs by tenure status. Finally, under a quality standard untenured professors will be awarded merit plus less often than tenured professors if and only if they are less productive. Yet in the data we find that untenured professors are on average as productive as tenured faculty members on measures we can observe, and yet they are awarded merit pay at significantly lower rates. Thus, the weight of the evidence indicates only a nominal adherence to a quality standard.

We identify two factors other than productivity which seem to influence decision-making. First, we find evidence of "warm glow" awarding in some departments. In these departments, a large majority of department members (including those not on the merit committee and therefore those with limited ability to engage in strategic behavior) are awarded merit plus. For these faculty members the hedonic value of deeming a colleague meritorious exceeds the monetary cost of doing so. A similar possibility is that decision-makers award merit plus to avoid backlash from unhappy colleagues.²

Second, we merge salary data with the university-wide data to investigate whether merit is used to address compression issues (i.e., to raise the salaries of senior faculty when their salaries are low relative to the salaries offered to new hires). We find some evidence for this but much of the variance in merit decisions remains unexplained.

These results empirically confirm the perception among faculty, as revealed in surveys (Quimby, Ross and Sanford, 2006) and in forums on Chronicle.com, that merit systems lack clarity and consistency. While the focus of this paper is positive rather than normative, we take a moment

² The Chronicle of Higher Education. 2004. "What Am I Worth?," April 23. http://chronicle.com/article/What-Am-I-Worth-/44570/

in the conclusion to list a few suggestions for how merit systems can be designed with greater transparency.

Given the inherent complexity of the issue and the immediate importance to a large number of economists, it surprises us that so little has been written in the economics literature about merit pay at universities. Much more attention has been devoted to the tenure system (e.g., McPherson and Schapiro, 1999; Carmichael, 1988; Dnes and Garoupa, 2005; Quimby, Ross, and Sanford, 2006). Like this paper, Euwals and Ward (2005) and Tuckman, Gapinksi, and Hagemann (1977) investigate the relationship between faculty remuneration and output. As expected, they find that research output positively influences a professor's salary. However, Euwals and Ward find that quality teaching is an important determinant of salaries while Tuckman, et. al. find only a weak positive relationship. This paper differs in two key respects. First, we have data on both annual merit decisions and productivity; the other papers do not observe raises directly. Second, we are able to identify annual changes in salary due to merit evaluations rather than salary level, which may be influenced by a variety of factors.

The paper proceeds as follows. In the next section we present a review of the literature, followed by our presentation in Section 3 of our theory of merit allocation which can be consistent with observed patterns of merit allocation. In Section 4 we describe the data and discuss evidence of warm glow awarding. Section 5 presents results from a more detailed investigation of the data and Section 6 concludes.

2. Literature review

The rewarding and retention of good faculty members is a priority of every school, and merit pay is one way schools may strive to do so. Studies have demonstrated the importance of quality teaching, showing that effective teachers can even compensate for the deficits experienced by children from disadvantaged backgrounds (Hanushek 2003, Rockoff 2004). Some studies show that merit pay can motivate above average performers (Marsden French and Kubo 2001) and that it can even be a more effective means of improving schools than upgrading equipment or facilities (Lavy 2002). However, even its supporters note that design of the mechanism is key, as a plentitude of pitfalls can render merit pay moot or even counter-productive.

A number of analysts conclude that merit pay is difficult to organize effectively in an educational setting. In a review of the literature, Hanushek (1986) finds that school expenditures are not linked to school performance, and merit pay in particular has been often tried but rarely persists. Burgess and Ratto (2003) note that early in their careers, workers need to demonstrate that they are hard workers, so additional incentives are redundant. Dixit (2002) notes that in the context of education, many outcomes are unobservable, and that measuring progress toward these outcomes is still harder. He concludes that "We should not expect [education] to turn into a[n]... organization that is left free to devise its own best procedures and judged by outcomes" (p. 721).

Incentives in such a context are tricky indeed. Instead of responding by increasing effort, government workers facing incentive schemes tend to "game" the system, sometimes to the detriment of the productive activity (Courty and Marschke 2003, Courty and Marschke 2004). Glewwe, Ilias, and Kremer (2002) describe a study in Kenya in which compensating teachers for student achievement achieved only fleeting gains, as teachers failed to even increase their own classroom attendance, instead simply shifting existing instruction into test-specific preparation.

Finally, some authors conclude that the offering of merit pay can actually be counterproductive, providing a disincentive to share information and function as a team as well as detracting from intrinsic motivation to work (Burgess et al. 2001, Belfield and Heywood 2008, Hanshaw 2004). If administrators or colleagues from similar disciplines are asked to do the evaluating, the process devolves into simple politicking. In fact, the larger the share of compensation taken up by merit pay, the more effort may be shifted from education to currying favor with one's evaluators (Adnett 2003). Further, merit pay skews incentives such that the appearance of yearly results supercedes risk-taking or long-term investment (Foldesi 1996). The

8

inability to adequately measure performance can become frustrating and sap teachers' motivation (Marsden French and Kubo 2001).

A first look at the data and now the literature both seem to cast aspersions on the usefulness of merit pay. Our theory section, however, shows that self-interested faculty members can successfully reward effort and do it in a way that is consistent with the observed data.

3. A Theory of Merit Allocation

Consider a department tasked with allocating a pool of merit money π among its faculty. Each faculty member is assigned either one merit point (*base merit*) or two merit points (*merit plus*). The value of a merit point equals the total value of the merit pool divided by the total number of points awarded. Thus, if a department has N members and $n \le N$ members are awarded merit plus, the value of a merit point is $\pi/(N+n)$. Those receiving merit plus get a merit raise of $2\pi/(N+n)$ while those receiving base merit get a merit raise of $\pi/(N+n)$.

There are two crucial observations about this merit system which affect incentives. First, the value of a merit point declines as more individuals are awarded merit plus. Thus, keeping one's merit level fixed, an individual prefers that fewer faculty members in the department receive merit plus (*MP*). Second, one is always at least as well off receiving *MP* compared to base merit (*BM*) regardless of how many others receive *MP*: $2\pi/(N+n) \ge \pi/(N+n')$ for any $0 \le n$, $n' \le N$, with equality iff n=N and n'=0.

Assume that all faculty members in a department can be distinctly ranked by quality based on observable output like publications and teaching evaluations. (Later we allow for ties.) When merit is allocated using a quality standard, all faculty members at or above a given position in the quality ranking are awarded *MP* and those below are awarded BM.³ The cutoff quality rank is known as the *quality standard*.⁴

Given the incentives inherent in the merit system, the quality standard that a purely selfinterested faculty member prefers is the one equal to her position in the quality rankings. This standard maximizes the value of a merit point under the constraint that she receives *MP*. Thus if one person, such as the department chair, has sole authority to select the quality standard, she can maximize her own return by selecting the quality standard equal to her position in the quality rankings.

In most departments, however, merit decisions are made by a merit committee composed of a subset of the department's members. (In many departments this committee consists of all tenured faculty members and no one else so we will sometimes refer to the committee as the tenured faculty). Note that a department with N faculty members will have to select from among N+1 quality standards since the department may always choose to award merit plus to no one, the *no-MP* standard.

Interestingly, in general there is no quality standard which is a Condorcet winner, where the Condorcet winner is the quality standard that wins by majority (sincere) vote in every pairwise contest. Consider the following example. The committee has three members, Ann, Bob, and Chuck. Ann is ranked highest, Bob second and Chuck third. Table 2 shows each member's preference ranking over the available quality standards where the number one indicates a person's most preferred quality standard.

³ Some departments have detailed written standards for merit plus. For these departments we can think of this model as a model of how the department agrees upon and updates these standards. For departments where the standards are vague, this model should be thought of as the annual merit allocation process.

⁴ In this model of merit allocation, it not necessary to assume that departments literally select a quality standard, but rather that most departments behave as if they do most of the time.

Quality standard	Ann	Bob	Chuck
No-MP	3	2	1
Α	1	3	2
В	2	1	3
С	3	2	1

Table 2: Preference Rankings over Quality Standards

The committee's preferences over the standards are intransitive under majority voting even though each committee member's preferences are transitive. This is a classic example of the Condorcet paradox. In fact, one can easily verify that this example is also a case of McKelvey's Theorem (McKelvey, 1976). That is, depending on the voting order, any quality standard may be chosen if the committee proceeds through the options by pairwise majority voting. Moreover, two common resolutions to the non-existence of a Condorcet winner fail to select a winner in this application, or to even reduce the number of quality standards the committee might select. Each standard has the same Borda count, and the Smith set, which is the smallest non-empty set of quality standards that will win against every standard outside the set, is {*No-MP*, *A*, *B*, *C*}.⁵

Given the weak power of standard voting solutions in this application, we require strong assumptions to narrow down the quality standards the committee may select. Assume the committee proceeds through the available quality standards by sincere pairwise majority voting, where the first contest is between the two highest standards and each subsequent contest pits the previous contest's winner against the next highest available standard. Conversations with professors across the campus under study lead us to believe that this voting order is a reasonable assumption.⁶ Fortunately, this assumption is required only to explain why such a large share of

⁵ More sophisticated resolutions such as those mentioned in Persson and Tabellini (2000) fail to narrow down the possible outcomes as well.

⁶ In practice the process for each department differs slightly. What is common to seemingly all departments is that a general discussion, either in a group or one-on-one, takes place after reviewing annual reports and before any voting begins. A ranking emerges. The person most deserving of merit plus is identified and then less qualified candidates

tenured faculty members receive merit plus. Under *any* quality standard, untenured professors receive merit plus less often if and only if they are less productive than tenured faculty.

With this voting order, a committee with an odd number of members will select the quality standard equal to the median committee member's quality rank, where the *median committee member* is the member whose quality rank has the property that at least half of the committee members are of equal or higher quality and at least half of the committee members are of equal or lesser quality.⁷

Proposition 1: Let $N \ge 2$. If there is an odd number of committee members, sincere pairwise majority voting selects the quality standard equal to the median committee member's quality rank when standards are considered in decreasing order.

Proof. Sincere voting means that committee members vote for whichever quality standard they prefer in every pairwise vote. In this case the *no-MP* quality standard wins every round in which it is paired against a quality standard that is strictly above the median committee member's quality rank. This is because every member at or below the median committee member's quality rank (a majority) strictly prefers the *no-MP* standard since he or she will receive *BM* under either standard and the no-*MP* standard implies a higher value of a merit point.

When the median committee member's quality rank is put up for a vote against the *no-MP* standard, the former wins since the members at or above the median committee member's quality rank constitute a majority, and these people prefer the former standard because they receive *MP*. This majority will hold as the committee proceeds through the remaining contests since lower standards reduce the value of a merit point and members in the majority are already getting *MP*.

are considered. In this sense the committee begins with the highest quality standard and then considers others in decreasing order.

⁷ A result similar to Proposition 1 arises for even numbered committees but this case introduces uninteresting details. These details are available upon request.

It is easy to allow for ties in this framework, and doing so will be important to explain why more than half of the tenured faculty receives MP. In this case the committee will have to select from anywhere between two and N+1 quality standards. There are two quality standards if everyone is the same quality (*no-MP* or everyone receives MP). As long as the median committee member is not also the lowest quality ranking in the *department*, Proposition 1 applies and the proof is identical. If the median committee member is also the lowest quality ranking to beat all standards strictly above the median committee member's quality rank, but now committee members are indifferent between the *no-MP* quality standard, in which case no one receives MP, and the standard equal to the median committee selects the latter quality standard in this case.⁸

This model explains why merit is broadly awarded in academic settings. Proposition 1 implies that at least half of the committee members (i.e., tenured faculty) will receive *MP*; an even larger share will receive *MP* if and only if there are ties at the median committee member's quality. Any systematic difference in merit awards between tenured and untenured faculty must be attributable to productivity differences. Thus, according to this theory the observation in Table 1 that untenured faculty receive merit plus less often than tenured faculty must be due to their lower productivity.

The reader can easily verify Proposition 1 in the example presented in Table 2 above. We recognize that even in this simple example committee members may improve their outcome by voting strategically when others vote sincerely. However, this model intentionally rules out such behavior to show how *MP* can be awarded to a large majority of tenured professors absent strategic considerations.

⁸ This assumption is justified if there is an infinitesimal hedonic value associated with receiving merit plus rather than base merit.

We can make one prediction from this model without productivity data. Since a larger committee requires more people to obtain a majority, when we control for department size we should observe a positive correlation between the number, but not the percent, of committee members receiving *MP* and the number of committee members.

With productivity data, we can test the fundamental assumption of this model that merit is allocated purely with regard to a quality standard. Either a person is above the standard or she is not. If this is true then we should expect that (i) the probability of getting *MP* is increasing in observable output like teaching evaluations and peer-reviewed journal articles, (ii) a large fraction of the variance in merit decisions can be explained by variation in observable output, and (iii) the effect of observable output is independent of tenure status. It is important to note that these three predictions are valid *independent* of our assumption about the order in which quality standards are considered by the committee.

3. The Data

To investigate the claims of the theory, we gathered three datasets from a large public university with over 500 faculty members.

3.1 University Level Data

The first set of data is an anonymous university-wide panel with 1587 observations on individual faculty in a total of 43 departments observed over one or more years. Each observation contains the final merit decisions during the most recent three-year period- 2007-2009, where merit decisions each year reflect activity the previous year (i.e. 2009 awards are based on academic year 2007-2008). This data set contains only a few variables for each observation: year, college, department, tenure status, and merit status (no merit, merit, or merit plus). See Table 3 for a summary. As we are interested in department level decisions, many of our tables are at the "department-year" level, which is the set of merit decisions made by a departmental committee in one year. Since some department-years contain few observations, we limit our sample to those with five or more faculty members, trimming the number of department-years to 100 from 117 and reducing the number of

faculty included to 1527.⁹ The unit of observation in these tables is the department-year, so the average "percent tenured" is the share of tenured faculty members in the mean department-year. The last four columns similarly reflect means of statistics calculated by department-year.

Award	# faculty	#	Avg %	Avg %	Avg %	Avg %	Avg %
Year	in sample	depts	tenured	merit plus	tenured MP	untenured MP	untenured MP*
2007	476	32	64%	64%	76%	57%	64%
2008	503	33	60%	68%	79%	56%	59%
2009	548	35	56%	66%	79%	54%	58%

Table 3: Sample Sta	atistics for the Trimmed	l Full Universit	y Sample ((N=100)
----------------------------	--------------------------	------------------	------------	---------

MP = Merit Plus. *Last column from untrimmed sample (N=117).

In the mean department-year, over 75% of tenured faculty are awarded merit plus. The only characteristic on which we found differences between the full and restricted dataset is the share of untenured professors receiving merit plus designation. Small departments (with fewer than 5 faculty members) award merit pay to untenured faculty at higher rates, an observation which kicks off our exploratory analysis section. Here we identify a few covariates associated with awarding merit plus in the university level data before next moving on to evaluating our theoretical predictions.

3.1.1 Department Size and Merit Plus

Before we look at behavior in departments of different sizes we must recall one characteristic of the award mechanism. Each share of the merit pay pool is a larger piece of the pie in small departments than it is in larger departments, where raising one member from merit to merit plus may diminish only slightly others' rewards. For example, if a department of 20 has \$20,000 to award for merit and they award everyone Base Merit, each will receive a \$1000 bonus. If they award just one person Merit Plus, then the value of a merit point falls to 20,000/21 = 952. This means 19 faculty members lose \$48 each by giving that one Merit Plus designation. If a department of 5 has a merit

⁹ We repeated all analyses of this dataset including departments with fewer than 5 faculty members as well as one department which had been subject to punitive action regarding its merit pay system, obfuscating the departmental decision process. This repetition helps us avoid drawing mistaken conclusions based on cases in which just one or two faculty members carry disproportionate weight in voting. In each case (except one, described below) we find no significant differences between a trimmed dataset and the full data, so we report only the former.

pool of \$5000, and they award one faculty member Merit Plus, then the value of a merit point falls to \$833. Thus four faculty members lose \$167 by awarding one Merit Plus designation.

Tables 4 and 5 show merit awards in small and large departments. In spite of the economic incentives to the contrary, in departments of four or fewer members all professors and specifically untenured professors seem to be slightly more likely to receive merit plus. The difference is significant at the 1% level. It may be politically more difficult to keep *MP* for the tenured in small departments since the cost of having an alienated department member may exceed the monetary compensation of the extra merit pay.

Award	# faculty in	#	Avg %	Avg %
year	sample	depts	tenured MP	untenured
_	_	_		MP
2007	16	6	90%	100%
2008	16	6	75%	80%
2009	15	5	100%	85%

 Table 4: Merit Awards in Small Departments (departments with < 5 total faculty members) (N = 17)</th>

Working in the opposite direction (i.e. first identifying departments awarding merit plus to a large share of the faculty and then checking department sizes) yields a similar conclusion (Table 5). First we identify the departments that are more liberal with their awards. The most generous quintile of departments bestows *MP* on an average of 96% of their faculty. Looking at department-years in this top quintile, we again find a highly significant (P(t) < 2%) association with department size. These departments, which tend to be smaller, appear to place more value on the hedonic importance of the designation than on its monetary value, and are hereafter referred to as *Warm Glow* departments. In these cases, the incentive function of the awards is basically nil.

Table 5: Warm Glow and Department Size

	Dept	#	% tenured	% MP	% MP
	size	untenured		among	among
				tenured	untenured
Warm Glow $(N = 22)$	9.8	3.4	55%	98%	89%
Regular ($N = 95$)	14.3	5.9	57%	74%	44%

3.1.2 College culture and merit plus

We further investigate heterogeneity in awards of merit pay by looking for different types of behavior in different colleges. Summary statistics by college are found in Table 6.

Table 6: Merit Pay by College

	College	College	College	College	College	College
	А	В	С	D	Е	F
% Merit Plus	52%	84%	52%	77%	69%	77%
Department size	11.2	14.1	14.3	8.2	22.9	11.1
% Tenured	68%	63%	62%	59%	62%	32%
% Tenured getting MP	59%	94%	63%	89%	79%	91%
% Untenured getting MP	36%	68%	36%	58%	53%	65%
Number dept-years in	14	18	30	18	15	22
data						

We see a fair degree of variation across colleges, and we have no information about particulars of each college's situation in the years we observe them, so we are at a loss for a robust explanation. Nonetheless, college level effects are apparent. For example as shown in Table 7 Warm Glow departments are predominantly found in just a few colleges.

Table 7: Warm Glow by College

	College	College	College	College	College	College
	А	В	С	D	Е	F
Warm Glow $(N = 22)$	0	8	1	5	0	8
Regular (N = 95)	14	10	29	13	15	14
% Generous	0%	44%	3%	28%	0%	36%

Warm Glow departments are found most frequently in College B and College F. Forty-four percent of the departments in College B are in the top quintile for wide distribution of merit plus, as are 36% of the departments in College F. Not far behind is College D, in which 29% of departments deemed all or almost all faculty worthy of the MP distinction.

We investigated whether the department size and the college level Warm Glow effects were actually the same, and found that they are quite distinct. In fact, College B has no departments with fewer than 5 members. Of 17 small department-years, 4 were in College C, 6 in College D, and 7 in College F. Regressing a Warm Glow dummy on department size and college dummies finds that department size has a negative coefficient that is significant at the 1% level, while the College B indicator is positive and significant at the 5% level with a very large coefficient.

The other side of the coin are the colleges where merit plus is awarded more parsimoniously. The bottom quintile of department-years awarded MP to an average of fewer than 36% of their faculty members. They are distributed as shown in Table 8. College A and College C are the most sparing with their awards, maximizing the individual value of the MP designation by awarding it to few faculty each year.

Table 8: Parsimony by College

	College A	College B	College C	College D	College E	College F
Parsimonious Dept-	6	1	15	1	1	4
Year $(N = 28)$						
Regular (N $=$ 89)	8	17	15	17	14	18
Share Parsimonious	43%	6%	50%	6%	7%	18%
(%)						

3.2 College Level Data

Our second dataset contains anonymous information for College A from 2005 - 2009. It includes details about the productivity of individual faculty members, including their teaching evaluations and publications during the year in question, as well as merit status. Teaching evaluations are filled out by students on one of the last days of the semester. Students are asked for their "Overall perception of the instructor," and they can respond with "Poor," "Fair," "Satisfactory," "Good," or "Excellent." These five choices are translated into 1, 2, 3, 4, and 5, and the mean of student responses is taken, yielding a number between 1 and 5. Publications are limited to peer-reviewed journal publications.

Since merit evaluation for an academic year is done at the start of the following year, first year faculty members are part of the department as of the time of evaluation but have no track record for the preceding year, so they are always awarded base merit and we drop them from the data. After dropping first year faculty members, the data set contains 189 faculty-years in five departments from 2005-2008. Twenty of these faculty-years lack observations on teaching evaluations (presumably because these faculty did not teach during the year in question). Three observations have data missing on the number of peer-reviewed journal publications and two more lack information on merit status, leaving us with 164 total. Summary statistics by tenure status for faculty-years including all relevant variables are shown in Table 9.

19

Table 9: Micro-data on faculty-years in College A

	N	Mean teaching evaluation score	Peer- reviewed journals	Merit award
Untenured Professors	34	4.27	0.85	44%
Tenured Professors	130	4.22	0.75	58%

T-tests, both raw and clustered at the department level, show that only the last difference is statistically significant at the 10% level.

Although considerable variation can underlie means, a first noteworthy fact from this data summary is that untenured professors do not immediately appear to be less qualified or less deserving of MP than tenured professors. In fact, by both criteria, they appear comparable to and even more productive than tenured faculty, though the difference is not statistically significant. However, they are less likely to be awarded merit plus.

While College A is not representative of MP allocation across the campus, it provides our most detailed glimpse into the process. Further, as one of the most parsimoniously awarding colleges, this dataset is particularly useful as it contains a high degree of heterogeneity in outcomes. Also, internal promotion and tenure documents provided to us by this college assert that teaching evaluations and publications are of primary concern in evaluating faculty.

3.3 University-wide Salary Data

Information on the salaries of employees at this university is public information. We acquired salary data for all employees as of September 2008 and computed departmental medians for tenured and tenure track faculty. These medians were used to compute the salary ratio of tenured faculty to untenured faculty by college as shown in Table 10. The university-wide merit pay data cover three years ending months before the period of time described by this salary data. However, they are the only salary data available, and salaries are unlikely to vary greatly from year to year.

Table 10: Median Salary Ratio of Tenured to Untenured by College

	College	College	College	College D	College	College F
	А	В	С		Е	
Tenured/Untenured	1.19	1.44	1.44	1.44	1.43	1.34
Ratio						

4. Results

4.1 Theoretical Predictions

What evidence is there for the use of a quality standard? The first specific prediction of the theory is that we should observe a positive correlation between the number, but not the percent, of tenured faculty members receiving merit plus and the number of tenured faculty members. In the university-wide data the correlation between the number of tenured faculty members and the *percent* of tenured faculty members receiving merit plus is about -1%. In other words, having a department with more tenured faculty members means that more people are deemed meritorious, but it does not mean that a larger share of tenured faculty members receive the designation.

To investigate the quality standard issue further we turn to data from College A which contains both merit decisions and productivity measures. As described above, the quality standard can be consistent with a large share of tenured faculty receiving merit plus. At least half should get MP, and ties could easily drive up the share achieving the award, particularly in departments where publications are the main criterion for merit status. Since publications are discrete, i.e. a person will be recognized as having zero, one, or two rather than 0.6 or 1.7, ties are likely to be common. For example, consider a department of five people for whom publications are the most important criterion for merit designation. If two people published articles and three did not, the median standard would be to give everyone merit plus.

21

The issue of ties makes sense when publications are paramount, but it is less appealing as an explanation when teaching evaluations, a continuous variable, are also considered. A quick look at our microdata from College A shows little clear demarcation between those receiving MP and those not receiving *MP*. Figure 1 shows that among tenured professors, one threshold is apparent at the three publication level, at which point all faculty receive *MP*. While the existence of this threshold lends support to the idea of a quality standard, it must be noted that this standard only applies in 7.5% of cases: just 10 of 130 tenured faculty-years reached this level of production. Although the existence of ties would enable us to explain the preponderance of *MP* designation among tenured faculty, the lack of a clear threshold in Figure 1 casts doubt on this possibility.





Figure 2: Merit Plus by Publications and Teaching Evaluation Scores, Untenured Faculty Only



Figure 2 illustrates the same schematic for untenured professors, and the absence of a threshold here is even more striking. Three of the five faculty-years in which a professor produced three or more peer-reviewed journal articles were deemed unworthy of *MP*. An appealing cluster of *MP*-receiving work is noticeable at teaching evaluations above 4.5 at the level of zero peer-reviewed journal publications, with a similar cluster between about 4.25 and 4.6. However, these are bounded on both sides by faculty deemed unworthy of *MP*.

As a first formal test of the data, Table 11 shows results from regressing merit status on teaching evaluation scores and peer-reviewed journal publications. We find that journal publications are significant at the 5% level, while evaluation scores are significant at the 1% level. Both have the expected positive signs. When we include a dummy variable for tenure, we find that the first two variables retain their signs and levels of significance, and that the tenure variable is positive and significant at the 10% level.¹⁰ If a quality standard is in effect, we would not expect to see the tenure variable coming in as significant, so this is a strike against the theory. However, its significance is marginal, so we still cannot draw a strong conclusion.

In the second half of Table 11, we split the effects of peer-reviewed journal articles and teaching evaluations into different variables for tenured and untenured faculty by multiplying the two variables by the tenure dummy. While peer-reviewed journal articles are positively and significantly associated with merit status when all faculty are grouped, the results are much weaker for untenured professors alone. The point estimate is less than one third the size of the coefficient on senior faculty, and it is statistically indistinguishable from zero. In other words, publishing in a peer-reviewed journal is not associated with an increase in the probability of obtaining merit pay for untenured faculty, though the association is present for tenured faculty. The story is slightly less stark in the case of evaluations, where the point estimate for untenured professors remains very close in size to the point estimate for tenured faculty. However, the standard error doubles, removing the statistical significance of the result. Higher teaching evaluations are associated with an increased likelihood of merit pay for tenured faculty, but the relationship is less clear for junior faculty.

¹⁰ We also include a department size dummy variable to account for the link observed above, and find that while it is not significant at traditional levels, it has the expected sign and its presence or absence has little effect on the other estimated coefficients. Similarly, we tested the effects of including fixed and random effects at the department level, but they were never significant and we dropped them. The remaining coefficients were almost completely unaltered by the presence or absence of these dummy variables.

	(1)	(2)	(3)	(4)
Peer-reviewed journal	0.11***	0.12***		
articles (PRJ)	(0.04)	(0.04)		
PRJ: untenured			0.06	0.06
only			(0.06)	(0.06)
PRJ: tenured only			0.16***	0.16***
			(0.06)	(0.06)
Evaluation score	0.33***	0.34***	0.35***	
	(0.10)	(0.10)	(0.10)	
Evals: untenured				0.27
only				(0.20)
Evals: tenured only				0.37***
				(0.11)
Tenured		0.18*	0.11	-0.29
		(0.10)	(0.12)	(0.80)
Pseudo-R ²	0.10	0.11	0.12	0.12

 Table 11. Micro-data Probit Regression Results¹¹

* significant at 10% level; ** at 5%; *** at 1% level. N = 164 faculty years for all regressions.

Interestingly, there seems to be a quality standard for tenured faculty but less of one for untenured faculty. This speaks against the idea of uniform application of a quality standard, but an obvious concern is the small sample size of untenured faculty-years (N = 34), which we unfortunately cannot remedy. However, the results are robust to trimming various sets of outliers. Four faculty-years were junior faculty who published 3 peer reviewed journal articles but were nonetheless denied merit plus designation. Dropping these observations (just over 12% of our sample of untenured faculty-years) raises the coefficient on peer-reviewed journals for untenured faculty, but still fails to generate a statistically significant result. Dropping all observations with more than two publications gives the same result. The conclusion that there is more heterogeneity in the link between publishing and merit pay of untenured professors is somewhat robust though the small sample size keeps us from making sweeping pronouncements.

More convincing to us are the results from the first column. While both publications and teaching evaluations are positively and significantly associated with *MP* status, together they

¹¹ This shows the results of four separate probit regressions with merit status as the dependent variable. In each box are the marginal effects coefficients with the standard errors in parentheses. All regressions include department size, which is always negatively signed but statistically insignificant (generally significant at the 15% level). Results change very little when OLS regression is used instead, or when department size is left out. Results are also robust to the inclusion of department dummy variables, which are insignificant.

explain a relatively low share of variation in *MP* status attribution. One hundred sixty-four observations are still not as many as we would like, but the fact that our set of explanatory variables is only sufficient to explain 10% of the variation in MP designation remains troubling. It seems more indicative of nominal adherence to a quality standard rather than of thoroughgoing implementation of such a standard. One of many possible explanations is that nominal adherence to some standard may be due to the presence of an outside monitor with the power to overturn decisions if they deviate too far from a quality standard.

Another concern is that we are so far unable to capture faculty service contributions, such as committee work or even serving as department chair. This unobservable is likely correlated with tenure status, biasing our results for tenured faculty upward. However, both anecdotal evidence and official documentation suggest that service work is weighed much less than teaching and research, so the impact is likely to be limited. We are confident that adding service indicators would not account for the other 90% of variation in MP assessment.

4.1.1 Tenured vs. Untenured MP

If untenured professors are as qualified as tenured professors and yet are awarded MP less, this is clear evidence against the quality standard. Table 9 shows that untenured faculty are on average just as productive as tenured faculty, but in fact they are awarded MP at much lower rates. No untenured faculty at all are awarded merit plus in 12-13% of both the trimmed and full samples, and in 20 of 110 department-years, tenured faculty received MP at a rate at least 50% higher than untenured faculty. Of the departments in which no untenured professors were granted merit plus status, just under half (7 of 15) awarded merit plus to 100% of tenured faculty. Overall the mean difference between the shares of tenured and untenured faculty deemed meritorious is about 28%.

We also investigated whether college or department size was a factor in predicting higher rates of MP among tenured vs. untenured faculty. Table 12 shows that department size does not appear to be correlated with the tendency to award MP to tenured faculty at a higher rate.

Table 12: Correlation of MP and Department Size

	Dept size	# untenured	# tenured
High differential $(N = 20)$	13.6	5.2	8.4
Regular (N $=$ 90)	14.3	5.8	8.5

Seven observations are lost in department years with either no tenured or untenured faculty.

Table 13 points to Colleges C and D as showing a slightly larger distinction between tenured and untenured faculty in the awarding of merit pay, but there is not much heterogeneity overall.

Table 13: Differential in rates of MP: Tenured vs. Untenured Faculty, by College

	College	College	College	College	College	College
	А	В	С	D	E	F
MP Differential	23%	26%	27%	31%	26%	26%

4.2 Is Merit Pay Used to Address Compression Issues?

If a quality standard is of limited use in explaining *MP* decisions, are there other variables systematically related to *MP* decisions? In this section we examine whether salary compression could be such a variable.

Salaries can and often do increase faster in the market than they do within a university. This can lead to a situation in which junior faculty members' salaries are close to or higher than that of more senior faculty members. This condition is called salary compression, and unsurprisingly it is frustrating to senior faculty. It seems possible that tenured faculty members may have the incentive to address compression through merit pay. If the merit committee acts on this incentive, we expect to see a low share of untenured professors getting merit designations, and more specifically, we expect to see a lower share of untenured professors getting merit designations in situations where compression is apparent.

Using our salary data with our full university data, we generated an index of compression in each department by dividing the median salary of full professors in the department by the median salary of untenured professors in the department (college level compression ratios are shown in Table 10). Unfortunately the salary data do not overlap exactly with the years of our full university data, but since the number of faculty are on the rise through the time period described by the merit pay data, new hires exceed the number of faculty leaving and compression too is likely on the rise. Thus, assessing compression at the end of the time period should identify the most egregious cases of compression. If *MP* has been used to combat compression in the past, that would work against our finding anything in the present data, as compression would be less prevalent thanks to the salary-increasing effect of *MP*. However, this is not what we see.

Compression ratios are available for 37 departments and range from 1.10 to 2.09. As reported in Table 14, we regressed the share of tenured faculty members receiving merit plus against a dummy variable indicating the top quintile of salary-compressed departments, and came up with a coefficient with the wrong sign and statistically indistinguishable from zero. We repeated the analysis including college level fixed effects and this time the coefficient was of the appropriate sign and significant at the 5% level. We then created a new dependent variable by dividing the share of tenured faculty receiving *MP* by the share of untenured faculty receiving MP, and regressed this upon the indicator for the most compression. The indicator was significant at the 10% level when it was alone, and when we included college and year effects, the size of the coefficient nearly doubled and the statistical significance improved to the 2% level. It seems that a larger degree of compression is associated with increased awarding of *MP* to tenured versus untenured faculty, casting doubt on the hypothesized reliance on a quality standard.

Dependent variable:	% of tenured	% of tenured	Share of tenured MP/	Share of tenured MP/
	getting MP	getting MP	Share untenured MP	Share untenured MP
Most compressed quintile	-0.08 (0.06)	0.16** (0.07)	1.35* (0.76)	2.64** (1.09)
Year indicators		Included		Included
College indicators		Included		Included
Ν	100	100	87	87
\mathbb{R}^2	0.02	0.41	0.04	0.10

Table 14: Regression Results including Compression Ratios¹²

** = significant at 5% level; * = significant at 10% level.

5. Conclusion

We conclude that while elements of a quality standard appear to motivate some of the merit decisions in the university, there is only weak evidence that this is the unique or even primary motivation. The need to satisfy overseers may contribute to this superficial appearance which is belied by largely unexplained heterogeneity. The difference in the rate at which tenured faculty are awarded merit plus relative to untenured faculty does not appear to be attributable to the superior qualifications of tenured faculty, even in a college where, without digging into the productivity data, there is weak evidence of differential treatment compared to other colleges. These observations are consistent with earlier work on merit pay in education by Adnett (2003), who notes that merit pay systems may encourage educators to devote their efforts to currying favor with evaluators rather than engaging in productive pursuits. Others such as Foldesi (1996) argue that merit pay distorts incentives with negative results for participant behavior as long-term investment and risk-taking is undermined by the need to meet yearly objectives.

Our theory section identifies a key tension in merit pay. Specifically, each faculty member desires a quality standard low enough so that he or she can be granted a high merit status, but also high enough so that the value of that merit status is not diluted. This can lead to the pattern of awards we observe, in which a majority of those with power to make decisions are awarded "merit plus" status. Curiously, however, likely common standards such as publications and teaching evaluations explain only a small part of the variation in awards.

¹² The mean (SD) of the first dependent variable is 0.77 (0.25), while that of the second dependent variable is 3.0 (3.0), so effects are relatively large in magnitude as well.

Instead, we find evidence that merit is used to address compression issues in departments where this issue is pronounced. In addition we observe some "warm glow" awarding, in which all or most faculty members in a department are said to warrant acknowledgement. This phenomenon is exceptionally common in small departments and in certain colleges, where like the children in Garrison Keillor's Lake Wobegon radio shows, everyone is above average.

We recognize that the preceding analysis is somewhat specialized to one university and constrained by small sample sizes for some of the analyses. Clearly the analysis could be improved with data from more universities. However, the advantage is that by focusing on one university we were able to become familiar with institutional details and idiosyncrasies that help us both in modeling the decision process and in the data analysis. Moreover, we suspect that many of the same incentives and issues that influence merit pay allocation at the university in this study are present at other universities. Finally, we suspect that a larger share of the variance in merit decisions could be explained with data on retention raises and outside offers. Unfortunately we do not have access to this data, but even if outside offers influence merit decisions, this would only reinforce this study's conclusion that the allocation of merit pay in practice deviates from the written policy on how merit should be allocated.

Despite a lack of clarity in merit systems and the debate regarding the effectiveness of merit pay for professors (c.f. Marsden, French, and Kubo 2001; Hanshaw 2004) and for educators more generally (e.g. Adnett 2003, Dee & Keys 2004, Lavy 2002, Podgursky & Springer 2006), it appears merit systems are here to stay, as stakeholders demand accountability from professors. In 1991 the president of one public university implemented a merit system in response to the state legislature's "discussions about the professor who was seen mowing his/her lawn at 2:00 in the afternoon instead of 'working'" (McMahon and Caret 1997). While untenured professors are highly incentivized by tenure, merit pay also provides a means for the university to incentivize tenured faculty.

30

So how can a merit policy be designed more efficiently? This is a question for future research but we provide some initial thoughts here. This paper identifies two crucial considerations. First, merit policy designers should strive to make the policy as transparent as possible on paper. Second, the policy should be designed so that those who implement the policy (merit committees) have little incentive or ability to deviate from its prescriptions.

In closing, we would again call the economic community's attention to the paucity of research in this area. Merit systems at many universities are in need of reform. Economists, particularly experts in mechanism design, are clearly well-equipped to make a significant contribution to this conversation.

References

Adnett, Nick. 2003. "Reforming teachers' pay: incentive payments, collegiate ethics, and UK policy." *Cambridge Journal of Economics* 27: 145-157.

Anderson, Kristine L. 1992. "Faculty Support for a Merit Pay System." Association for the Study of Higher Education Annual Meeting Paper. 38pp.

Belfield, Clive R., and John S. Heywood. 2008. "Performance Pay for Teachers: Determinants and Consequences." *Economics of Education Review* 27: 243-252.

Burgess, Simon, Bronwyn Croxson, Paul Gregg, and Carol Popper. 2001. "The Intricacies of the Relationship Between Pay and Performance for Teachers: Do Teachers Respond to Performance Related Pay Schemes?" CMPO Working Paper Series No. 01/35, July, 32pp.

Burgess, Simon, and Marisa Ratto. 2003. "The Role of Incentives in the Public Sector: Issues and Evidence." *Oxford Review of Economic Policy*, Summer 19(2): 285-300.

Carmichael, H. Lorne, 1988. "Incentives in Academics: Why is There Tenure?" *Journal of Political Economy* 96(3): 453-472.

Courty, Pascal, and Gerard Marschke. 2003. "Dynamics of Performance measurement systems." *Oxford Review of Economic Policy* Summer 19(2): 268-284.

Courty, Pascal, and Gerard Marschke. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22(1): 23-56.

Dee, Thomas S., and Benjamin J. Keys. 2004. "Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment." *Journal of Policy Analysis and Management* 23(3): 471-88.

Dnes, Antony and Nuno Garoupa, 2005. "Academic Tenure, Posttenure Effort, and Contractual Damages," *Economic Inquiry* 43(4): 831-839.

Dixit, Avinash. 2002. "Incentives and Organization in the Public Sector: An Interpretative Review." *Journal of Human Resources* 37(4): 696-727.

Elhagrasey, Galal M., J. Richard Harrison, and Rogene A. Buchholz. 1999. "Power and Pay: The Politics of CEO Compensation." *Journal of Management and Governance* 2: 309-332.

Euwals, Rob and Melanie E. Ward. 2005. "What Matters Most: Teaching or Research? Empirical Evidence on the Remuneration of British Academics." *Applied Economics* 37: 1655-1672.

Foldesi, Robert. 1996. "Higher Education Compensation Systems of the Future." *CUPA Journal* Summer 47(2): 29-32.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2002. "Teacher Incentives." Mimeo, Brookings Institution, Washington DC, November.

Gravois, John. 2007. "U. of Washington Settle Lawsuit Over Pay for Part-Time Lecturers." *Chronicle of Higher Education* July 13, vol. 53. 1p.

Hammond, Ron J., Pat Ormand, Terry Nichols, John Balden, Linda Edgeton, Keith Snedegar, Denza Bruss, Linda Makin, and Karl Worthington. 1999. "An Exercise in Growth and Adaptation for a Rapidly Growing State College: Faculty Pay Scale Report and Proposals. Presented to the Faculty Senate by the UVSC Faculty Senate Budget Committee, 1998-1999." Utah Valley State College, Orem, Utah.

Hanshaw, Larry G. 2004. "Value-Related Issues in a Departmental Merit Pay Plan." *The Professional Educator*. Spring XXVI(2): 57-68.

Hanushek, Eric. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*. September 24(3): 1141-1177.

Hanushek, Eric. 2003. "The Failure of Input-based Schooling Policies." *Economic Journal* 113 (February): F64-F98.

Lavy, Victor. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* December 110(6): 1286-1317.

Marsden, David, Stephen French, and Katsuyuki Kubo. 2001. "Does Performance Pay De-Motivate, and Does It Matter?" Center for Economic Performance/ LSE Working paper, August, 43pp.

McKelvey, Richard D. (1976). "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control." *Journal of Economic Theory* 12, 472-482.

McMahon, Joan D., and Robert L. Caret. 1997. "Redesigning the Faculty Roles and Rewards Structure." *Metropolitan Universities* Spring: 11-22.

McPherson, Michael S. and Morton Owen Schapiro. 1999. "Tenure Issues in Higher Education." *Journal of Economic Perspectives* 13(1): 85-98.

Persson, Torsten and Guido Tabellini, 2000. *Political Economics: Explaining Economic Policy*.MIT Press: Cambridge.

Podgursky, Michael J., and Matthew G. Springer. 2006. "Teacher Performance Pay: A Review." National Center on Performance Incentives Working Paper. October 24. 52 pp.

Quimby,J., Ross,D. & Sanford,D. "Faculty Perceptions of Clarity in Promotion and Tenure Evaluations" <u>The Teachers College Record</u> (Columbia University) August 16, 2006. <u>http://www.tcrecord.org</u> ID Number: 12668.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 94(2): 247-252.