The empirical verification of key linkages in the Transit System's Pay-for-Performances Program

S E Markham
Dow Scott, Loyola University Chicago
F S Hills

Available at: https://works.bepress.com/dow_scott/29/
THE EMPIRICAL VERIFICATION
OF KEY LINKAGES IN
A TRANSIT SYSTEM'S
PAY-FOR-PERFORMANCE PROGRAM

STEVEN E. MARKHAM
K. DOW SCOTT
FREDERICK S. HILLS
VirginiaPolytechnicInstituteandStateUniversity

Transit authorities provide direct service to the public and represent a major cost to local, state, and federal governments. [1] They receive close public scrutiny when resource dollars are allocated. In response to pressure for increased productivity and better quality of service, transit authorities are reexamining the use of financial incentive programs. While various types of financial incentive programs are commonly found in both the private and public sectors, merit pay programs are the most widely used. A survey conducted by the Conference Board (1976) found that over 90% of the responding companies used a merit pay program. A study by Scott and Deadrick (1984) found that merit pay is the most popular financial incentive program among transit authorities, with 30% of respondents indicating that they had a merit pay program at least for their office or managerial employees.

Merit pay programs are indeed common, but the ability of OD professionals to use them to support system-wide change efforts has been limited. Nevertheless, an important linkage exists between compensation systems and organizational development (OD) efforts (Lawler, 1981a). The need for structural rewards to support more process-oriented OD interventions will be an important theme for the future of OD (Golembiewski, 1986). This research can be viewed as a preliminary effort to develop a diagnostic framework which can be used by either the compensation analyst or the OD facilitator.
Many pay plans are designed to establish a linkage between an employee's salary and performance. They traditionally contain two components: (1) the potential for an individual merit raise and (2) an appraisal of individual performance. Merit raises are awarded based on employee performance over a specified time period (usually annually). An overall merit budget is established by top management and is usually based on (1) the past financial performance of the organization; (2) expected future performance; and (3) the necessity of adjusting pay schedules upward. The merit raise is designed to reward those employees who have performed at high levels during the past period. The merit award usually represents a permanent increase in the employee's wage or salary and thus a permanent increase in the organization's total labor costs.

Regardless of the many varieties of administering merit pay systems, inherent in any program is the need for some reliable method of appraising individual performance. This appraisal usually takes the form of a periodic evaluation by the employee's superior. Behaviorally-anchored rating scales (BARS), management by objectives (MBO), and graphic rating scales are the most commonly used appraisal instruments (Bernardin and Beatty, 1984).

While the widespread and continued use of merit pay plans suggests that managers believe such programs enhance employee performance and productivity, the value attributed to merit pay programs is largely an act of faith. There is very little empirical evidence with which to evaluate these programs (Dunnette and Bass, 1963; Haire et al., 1967; Lawler, 1971, 1981b). Furthermore, a number of experts have identified weaknesses in merit pay programs and contend that these programs can create motivational problems if the programs are improperly used (Meyer, 1975; Hills, 1979). This research operationalizes a framework for the evaluation of pay-for-performance programs in a transit environment so as to provide a means for responding to the above critiques.

Research on Merit Pay

It is a widely-held assumption of modern compensation
programs that a portion of an employee’s remuneration should be intimately tied to performance as a means of obtaining the desired work behaviors. Numerous researchers such as Lawler (1971) and Locke et al. (1980) have provided an impressive array of evidence showing that, where pay is related to individual performance, employee motivation is increased and turnover is largely restricted to poor performers. More specifically, Locke et al. (1980) found that financial incentive programs had a far larger impact on employee performance than did other major management programs (e.g., participative management, job enrichment, and goal setting).

A review of the literature revealed 94 articles on merit pay. However, the majority of these articles were simply descriptive or conceptual in nature. Given the widespread use of pay-for-performance programs and the diversity of goals for which they can be designed, it is somewhat surprising that only 16 articles empirically evaluated various aspects of these programs. In one of the earliest studies, Marriott (1962) focused on employee attitudes about merit pay plans. No employee was rated below average, yet most employees were very dissatisfied with the system. Were all employees rated too low? Was there some other reason for the dissatisfaction? While there is insufficient information available in the Marriott study to answer these questions, that research raises some of the common issues in the administration of pay-for-performance programs.

On a more positive note, Kopelman, Reinhardt, and Beer (1983), in a study of 1,165 nonsupervisory white-collar employees, found that having a wide range of rewards available increased overall performance. Strong ties between performance and rewards resulted in high achievement. Furthermore, Elliot (1985), in a study of 60 personnel professionals and public employees in federal, state, and local governments, found that there was general support for merit pay programs.

Building on these efforts, Hills, Madigan, Scott, and Markham (1987) have suggested a number of specific ways to evaluate a pay-for-performance system. These criteria are summarized as follows:

1. Is there a significant difference in the average merit raise
percentage awarded to low, medium, and high performers? While one would expect the answer to this question to be affirmative for any system, Markham (1988) has provided an example of a system where, at the individual level of analysis, there was a non-significant correlation between performance rating and merit raise. Thus, the degree to which performance rating and the merit award are related (or correlated) is a simple empirical descriptor which can vary from system to system.

2. Are there significant differences between men and women or between racial groups in terms of merit raises awarded? To the extent that there are major differences, then it is possible that the compensation system has failed to meet its goals in the affirmative action/equal opportunity arena.

3. Are there significant differences between supervisory groups? The problem of supervisory biases in ratings is well-known (Hills, 1987). If major differences exist between supervisory groups, it is questionable if employees will perceive equity in the system, regardless of the explanation for those differences.

4. Is there a "job level" effect on the size of merit pay increases? If larger merit pay increases are given to those employees who are in the upper levels of the hierarchy, then there may appear to be a bias in the compensation system towards those of higher status.

5. Do high performers remain with the organization? In other words, given the goal of retaining the most valuable employee, are those who leave the system primarily the low and medium-level performers? Clearly, retaining the best performers would be the ideal situation.

6. Are those who are promoted drawn from the ranks of the high performers? One use of a merit system is for promotion decisions. If those who are promoted are drawn equally from the low and middle levels of performers, one would have to question the usefulness of the performance rating data in selecting upper level managers.

7. Do those who receive high merit raises one year also receive them in future years? Hills et al. (1987) suggested that a year-to-year analysis should reveal a significant
correlation to demonstrate that a random process is not occurring in which those who are high one year are low the next year so as to "share the wealth" by rotating the raises.

8. Finally, do employee attitudes reflect satisfaction with their performance ratings and merit raises? One would certainly not want a pay system that created widespread dissatisfaction or distrust.

**METHODODOLOGY**

*Research Site*

The research site for this study is a large transit authority in the United States. This authority employs over 5,000 employees with over 1,000 participating in the transit system's merit pay plan. [2] This transit authority maintains a fleet of over 1,500 buses. During the fiscal year 1984-1985, it carried over 300,000,000 passengers. The researchers reviewed the pay systems of over 300 transit maintains to identify a research location that had a sound merit pay program. Based on their evaluation of policy statements and methods used to administer the program, they believed that this location's program was designed properly and was being administered effectively. Therefore, they were asked to participate in the study.

*Data Collection*

Information was collected from employee records on approximately 1,087 employees for 1983, about 1,189 employees for 1984, and 1,255 employees for 1985. The three-year history of employee performance appraisal scores and employee merit increases was supplemented by information concerning gender, supervisory group membership, and job level. The performance appraisal and the merit raise measures are explained separately below. Job level was a categorization used in the salary structure which indicated the relative difficulty of the work performed. It was available for 1983 and 1984 and ranged from 1 to 27 with a higher number indicating a more difficult, better-
paid job.

*The Performance Rating Measure.* The performance appraisal evaluation is a rather elaborate document comprised of separate supervisory ratings of the employee on work habits, task behaviors, and results. The supervisor is required to combine all of this information "clinically" into an overall judgment of performance. This performance evaluation is reviewed by the personnel department and all higher levels of management including the director of the transit authority. For this study, it is this overall evaluation which is used as the measure of performance level. The scaling was converted to a numeric system by assigning values of 1 through 5 to the performance levels, which included "unsatisfactory," "needs improvement," "competent," "superior," and "outstanding."

*The Merit Raise Measure.* Merit raise is the annual pay increase given to each employee. The pay increase becomes a permanent part of the base wage and is expressed as a percentage derived from the division of the current wage increase by the previous year's base wage. In 1981, the average wage increase was 7.8%, in 1984 5.5%, and in 1985 5.2%. Not all of these increases, however, were for merit awards. Some of each year's increase was a system-wide pay adjustment and some was for structural adjustments. These adjustments were for employees who fell at the bottom or below the pay range and had to be adjusted up to stay within the structure. To remove some of these adjustments and make easier comparisons across years, the merit increase data have been adjusted by subtracting out the year's average from each individual's increase. Thus, it is possible for some individuals to show a negative "increase" relative to the baseline of zero.

During the third year, approximately 800 employees completed a questionnaire (a 64% response rate) designed to tap their attitudes towards the merit pay program. The age of respondents ranged from 26 to 67 years with an average of 44.8 years. Respondents had an average length of service of 13.5 years with a range of 1 to 44 years. Seventy-eight percent of the respondents were male and 22% were female. A wide variety of ethnic backgrounds was represented: 26% were black, 55% were white, 7% were Oriental, and 13% were Hispanic. They
also had varied educational backgrounds: 8% had a high school diploma, 43% had some college, 21% had a college degree, 11% had some graduate work, and 15% had a graduate degree. Two specific items from this survey regarding satisfaction with the merit increase and satisfaction with the performance appraisal rating are reported in the following section. This questionnaire was pilot tested at a small transit authority on the East Coast which had a merit program.

A research team from the authors' university administered the questionnaire. Participants were notified by memoranda from the Personnel Department of the research project and were requested to attend one of the scheduled sessions to fill out the questionnaire. Individuals who did not attend one of the formal sessions were provided with questionnaires and instructions for returning them by mail. Confidentiality of information was stressed heavily in these sessions because employees were asked to put their names on the questionnaire so that their responses could be matched with employee records.

RESULTS

(Q1) Consistency Between Merit Raises and Performance Ratings. A one-way Analysis of Variance for unequal cell sizes was computed for performance appraisal scores using merit raise as the dependent variable. This was repeated three times for each of the years in question. There was a significant difference in merit raise between the levels of performance rating for each year. These significant differences accounted for 20% (F = 80, df = 3,938, p < .0001), 11% (F = 45, df = 3,1094, p < .0001, and 56% (F = 53, df 4,1251, p < .0001 respectively, of the variance in merit raise. As shown by the average raise scores in Table 1, the highest performers consistently received the highest raises. For example, in 1983 the outstanding performers (N = 43) received a raise that was 3.35% above the year's average of 7.8%. In 1984 the outstanding group (N = 68) received a 3.00% increase over the smaller baseline of 5.5%. However, there did not appear to be a significant difference, as noted in the Duncan's results, between the "superior" and "competent" groups which contributed to the relatively low
<table>
<thead>
<tr>
<th>Improvement</th>
<th>Needs</th>
<th>Complete</th>
<th>Superior</th>
<th>Outstanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>5.17%</td>
<td>4.34%</td>
<td>8.79%</td>
<td>(2)</td>
</tr>
<tr>
<td>C</td>
<td>6.05%</td>
<td>4.92%</td>
<td>9.14%</td>
<td>(3)</td>
</tr>
<tr>
<td>B</td>
<td>7.75%</td>
<td>6.09%</td>
<td>9.42%</td>
<td>(4)</td>
</tr>
<tr>
<td>A</td>
<td>9.40%</td>
<td>7.99%</td>
<td>9.34%</td>
<td>(5)</td>
</tr>
</tbody>
</table>

Table 1: Performance Rating Levels by Year Using Duncan's Test
R² term for that year. In 1985, as in 1983, each performance group's average raise was significantly different from all other groups as indicated by the Duncan's results, and there was less variance within each performance category as indicated by the much higher R² term of 56%

As a by-product of the above analysis, the authors noted that there was a "slow creep" over the three years in the performance appraisal scores. In 1983, 43% of the employees were evaluated as superior (the next to the top category) and only 4.6% were evaluated as outstanding (the top category). In 1985, 44.5% were rated as superior and 8.5% were rated as outstanding. Thus, the average performance rating increased from 3.46 (s.d. = .60) the first year and 3.50 (s.d. = .63) to 3.60 (s.d. = .67) during the last year. The last year's score was significantly higher than the scores of the previous two years (Duncan/Scheffe test, F = 16.69, p < .0001).

(Q2) Gender and Racial Group Differences. With respect to merit increase, there was no significant differences between the racial groups for the first two years. In 1985 the 60 Asians received an average increase which was significantly higher (F = 4.19, p < .006, R² = 1.3%) than the Hispanics (N = 115), Blacks (N = 278), and Caucasians (N = 498).

On the other hand, there were differences between men and women for all three years. In 1983 there was a statistically significant difference between the average net merit raise of -.97% below the baseline for men (N = 729) compared to the women's net average of .18% (N = 217, F = 20.1, p < .0001, R² = 1.7%). In 1984 the men's net raise averaged -.24% below the baseline and the women's averaged .46% above (F = 12.7, p < .0004, R² = 1.1%). Finally, in 1985 the average net raise for men was -.09% versus .12% for women (F = 5.5, p < .02, R² = .04%).

(Q3) Supervisory Group Differences. There were significant differences in the average merit raise given by supervisors to their employees for each of the three years in question. This analysis used a one-way ANOVA for unequal cell sizes with supervisory groups aligned as cells. A version of this procedure is described elsewhere (Markham, 1988).

In 1983 about 40% (F = 2.23, df = 208, 703, p < .0001) of the
variation in merit raise occurred between supervisory units. This figure was reduced the next year to 29% (F = 1.43, df = 237, 833, p<.0001) and it remained fairly stable at 27% (F = 1.32, df = 263, 930, p<.002) the last year. It was not uncommon to have entire units of 5 to 7 individuals who received raises 7% above the baseline in 1983. Similarly, there were entire supervisory units which were awarded "increases" of -.7% below the baseline for the year.

(Q4) The Job Level Effect. To investigate the question, a Pearson correlation was computed by year between an individual’s merit raise and his/her job level score. For 1983 the relationship was r = -.13 (p<.0001). For the last two years, the relationships were -.05 (p = .13) and .01 (p<.73), respectively.

(Q5) The Retention of High Performers. To address this question, the authors identified the employees who left the system after the 1984 merit raise. By going back to the previous year, the authors were then able to categorize employees either as Stayers or Leavers. Again doing the one-way ANOVA for unequal cell sizes, the significance of the difference by merit raise was calculated. In 1983, those who would leave (N = 68) during the next year had an average loss of -.22% in contrast with .02% for those who stayed (N 877). This was not a significant difference.

(Q6) The Promotion of High Performers. To address this question, the authors identified the employees who were promoted at least one job grade from 1983 to 1984 and used this factor to categorize employees as either Stable or Promoted. Again, using a one-way ANOVA, the significance of the difference in merit raise was calculated. In 1983, those who would be stable (N = 819) had an average raise of -.11% below the baseline. Those who would be promoted (N = 55) had an average raise of 1.17%. This was a significant difference (F = 7.5, p<.006, R² = 1%). As would be expected, after the promotion took effect the next year, there was a much larger difference in merit raise between those promoted (5.34% above the baseline) and those who were stable (-.33 below, F = 282, p<.0001, R² = 21%.

(Q7) Year-to-Year Comparisons. To answer this question, a Pearson correlation was computed using each employee’s merit
<table>
<thead>
<tr>
<th>Question:</th>
<th>SA</th>
<th>MA</th>
<th>?A</th>
<th>?D</th>
<th>MD</th>
<th>SD</th>
<th>na</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) I am very satisfied with the</td>
<td>99</td>
<td>96</td>
<td>67</td>
<td>44</td>
<td>39</td>
<td>121</td>
<td>1</td>
</tr>
<tr>
<td>last performance evaluation I</td>
<td>21%</td>
<td>21%</td>
<td>14%</td>
<td>9%</td>
<td>8%</td>
<td>26%</td>
<td>.2%</td>
</tr>
<tr>
<td>received.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) I am very satisfied with the</td>
<td>31</td>
<td>76</td>
<td>80</td>
<td>71</td>
<td>61</td>
<td>109</td>
<td>4</td>
</tr>
<tr>
<td>last merit increase I received.</td>
<td>7%</td>
<td>16%</td>
<td>17%</td>
<td>15%</td>
<td>13%</td>
<td>30%</td>
<td>.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1986 Correlations of Satisfaction with Merit Raise
Table 3

Item 1 and 3 were reversed for the correlation calculation.

Approximate N = 467

1000 > p***

1 2 3 4
x 72** 44*** 22**

x 65*** 39***

x 22***

x

(4) Actual Performance Rating 85
(3) Satisfaction with Performance Rating
(2) Actual Merit Raise 85
(1) Satisfaction with Merit Raise

And Performance Appraisal
raise in one year and his/her raise the following year. From 1983 to 1984, the correlation was $r = .17$ (p < .0001). From 1983 to 1985, the correlation was $r = .24$ (p < .0001). Finally, from 1984 to 1985, the correlation was $r = .39$ (p < .0001).

(Q8) **Employee Satisfaction with Merit and Appraisal.** A final way to evaluate a merit pay system is to solicit employee opinions about it. For this part of the study, the researchers asked employees a series of questions to ascertain their attitudes toward the merit pay system. A full discussion of all the attitudinal data associated with this project can be found in Hills, Scott, Markham, and Vest (1987). Two key measures, Satisfaction with Raise and Satisfaction with Performance Appraisal, are shown in Table 2 along with their frequency distribution. (For these questions, the abbreviation "SA" means Strongly Agree, "MA" means Moderately Agree, "?A" means Marginally Agree, "?D" means Marginally Disagree, "MD" means Moderately Disagree, and "SD" means Strongly Disagree.)

The correlation between the degree of satisfaction with the raise and the actual amount of the merit raise was $r = .40$ (p < .0001, N = 473, reverse coded) as shown in Table 3. The matching correlation between the level of satisfaction with the performance appraisal and the actual appraisal rating was $r = .67$ (p < .0001, N = 466, reverse coded).

These correlations should be approached with caution for two reasons. First, they use single-item measures. Second, the two attitudinal measures had unusual distributions. In fact, the most interesting observation about the results of these questions is that employees seemed to fall into two extreme groups of employees, as shown in Table 2. There was a substantial proportion (21%) who were highly satisfied with their last performance appraisal; however, there was a substantial proportion (26%) who were just as highly dissatisfied with their last performance appraisal. Clearly, employees were deeply divided in their attitudes toward satisfaction with their last performance appraisal. This effect was even more pronounced with respect to merit pay, with only 7% strongly agreeing with the statement that they were very satisfied with their raise versus 30% who were at the other end of the continuum.
DISCUSSION

The empirical "portrait" of this transit system's pay-for-performance program suggests a number of strengths and weaknesses. This is to be expected in that the merit raise program was introduced in 1983 and was relatively new when the study was conducted.

Q1: Are high performers rewarded? For each of the three years of data, the researchers found significant differences in the average raise percentage awarded to low, medium, and high performers. The relationship between the performance rating and the merit award may be viewed as a simple empirical descriptor which can vary from system to system. Here, for example, the performance ratings accounted for 20%, 11%, and 56% of the variation in merit pay. However, we cannot evaluate the adequacy of this relationship because we do not know how strong this relationship must be in order for employees to perceive a linkage between performance appraisals and merit increases. Nor do we know how other transit systems would compare on this very important criterion.

Q2: Are there significant differences between the sexes or racial groups? It appears that the merit system has supported its goals in EEO/AA. During the first two years of the program, there was no differences based on race. In the last year, only the Asian group was significantly higher. Also, while there was significant differences between the sexes, it was the women, not the men, who benefited by receiving higher wages. (This occurred because women were lower in the pay range and had less experience on average than men.) Previous historical data were not available to determine if inequities existed prior to 1983.

Q3: Are there significant differences between supervisory groups? This analysis reveals that all supervisors do not rate their employees the same. The data indicate that there are large differences between supervisors in the average performance score they give to their subordinates. Over the course of the three years, this problem became less pronounced. In 1983, about 43% of the variation in scores occurred between units. The next year, it was 29%, and in 1985 about 27% of the
variation in performance evaluation scores occurred between supervisors. The problem of major differences between supervisory units can be serious because a perception of unwarranted bias can jeopardize the perception of equal and fair treatment for all employees. It is possible that some supervisors are more lenient in their evaluations while other supervisors are excessively strict in their evaluations. In light of the levels of dissatisfaction with the merit pay system, the second scenario warrants investigation. Alternatively, the data could mean that there are true differences in average performance between supervisory units.

Q4: Is there a job level effect? The data suggest that, in the first year of the program, the higher one’s job level, the higher one’s merit raise. During the last two years of data, this effect disappeared.

Q5: Are the high performers retained in the system? In the year for which this could be fully calculated, there was not a significant difference in the raises given to those employees who would later leave the system. It is possible that those who left did so because of external economic consideration rather than poor performance.

Q6: Are those who are promoted drawn from the ranks of the high performers? For the year for which this could be fully calculated, it appears that those who were to be later promoted received a higher raise than those who were not promoted. The difference in the size of their raises was even more pronounced after the promotion. This finding suggests that the merit system is linked to the promotion decision process.

Q7: Do those who receive high merit raises one year also receive them in future years? A year-to-year analysis revealed a significant correlation between the first two years’ merit raises ($r = .17, p < .0001$). This relation improved during the last two years ($r = .39, p < .0001$). Apparently, a random process was not occurring in which those who were high one year were low the next year so as to “share the wealth” by rotating the raises. However, the researchers have no comparative data to suggest how high this relationship might be in other organizations. More important, the correlation of .39 means that about 85% of the variance was not accounted for by the previous year’s raise.
Thus, many other factors can help explain variation in merit raises.

Q8: Are employee levels of satisfaction linked with the size of their pay increase? Finally, two key employee attitudes reflected a nearly bimodal distribution of satisfaction with their performance ratings and merit raises. Those employees who got high ratings were very satisfied, while those who received low supervisory ratings tended to be dissatisfied. These results are suggestive of a number of possible scenarios. In one scenario, if one assumes that supervisory ratings are reflective of true performance, then the performance review system and the pay system are doing their job—rewarding high performers. The fact that some employees are dissatisfied is a natural result of their own failure to perform at high levels. On the other hand, in a second scenario in which supervisory performance scores are not accurate reflections of true performance, then there are some high performers who are not being fairly treated and there are some low performers who are being over-rewarded. As a result, the merit pay program may be creating unwarranted dissatisfaction which may be disruptive to the organization. Equally important, this dissatisfaction may well be reflected in the levels of stress and psychological burnout experienced by individual employees (Deckard, Rountree, and Golembiewski, 1986). Unfortunately, we have no comparative data from other transit organizations with which to put this into perspective.

CONCLUSION

Although this study does not test specific hypotheses with respect to motivation and pay increases, it does represent an in-depth examination of a pay-for-performance program in a large public sector organization. Such an examination should be part of any OD diagnosis prior to a system-wide change effort.

The authors purposely selected a research location that was believed to have a sound merit pay program. The amount of dissatisfaction with the pay system was surprising. Why this occurred is difficult to ascertain. Possibly the program or the administration of the program was flawed and employees were reacting to a poorly designed merit pay system. On the other
hand, employee perceptions of the program may have been distorted, given the inherent difficulties involved in the develop-
ment of a merit pay system.

From a longitudinal perspective, however, the system seemed to have made important improvements in terms of (1) increasing the strength of the pay-for-performance linkage, (2) the lack of gender/race problems, (3) decreasing the supervisory group effect, (4) decreasing the job level effect, and (5) signaling to and rewarding those who were promoted. In order to develop a comparative data base, further research is certainly needed. This research can be viewed as an initial step in developing criteria for comparing pay-for-performance programs across transit systems prior to instigating a general OD intervention.

NOTES

1. This study was funded by a grant from the Office of Technical Assistance of the Urban Mass Transportation Administration, U.S. Department of Transportation. The U.S. government assumes no responsibility for the contents of this article. Portions of this article were presented at the 1987 Academy of Management National Meetings.

2. These numbers are given as approximations in order to disguise the identity of the research location.

REFERENCES


