WORKER PERFORMANCE AND GROUP INCENTIVES: A CASE STUDY

DANIEL G. HANSEN*

The author uses monthly, individual-level data collected from two units within a large U.S. financial corporation to estimate employees' response to the introduction of a group incentive plan. The findings indicate that the incentive plan caused performance to converge to a standard: the initially least productive workers improved greatly, whereas the performance of the initially most productive workers did not change. However, the evidence suggests that the incentive plan was successful in increasing the average level of productivity across the work group.

It is well known that firms have difficulty getting the desired level of productivity out of their workers. Piece rate wages and individual incentive plans can lead to higher individual effort, but these systems present problems. First, they do not encourage cooperation. For example, a bonus that pays only to the worker with the highest sales volume puts the members of the sales force in competition with one another, providing an incentive to sabotage a co-worker's efforts. Second, it may not be possible to implement an individual incentive because the output of individual workers is often difficult to measure. Consider, for example, a team of workers that is responsible for developing a software application. When the application is produced, it may be difficult to determine what was contributed by any one worker.

An alternative is to offer incentives based on group productivity in the form of either profit sharing or gain sharing plans. A profit sharing plan simply shares a fraction of the profits the company earns with the workers, usually on an annual basis. Gain sharing plans are based on more specific goals, with workers sharing in the financial gain that comes from achieving these goals. Several variants of gain sharing have been used, some with very specific compensation formulas. For example, in the Rucker Plan, value added is multiplied by a historical ratio of payroll costs to the dollar value of production to compute the allowed labor costs, the actual labor cost is subtracted from this value, and what remains is shared between the workers and the managers.

Ideally, gain sharing plans are linked to quantifiable goals (such as fewer program-

*The author is Economist at Laurits R. Christensen Associates. He thanks Daniel Hamermesh, Harry Holzer, Robert LaLonde, the members of the Michigan State University Labor Lunch Group, Robert Hutchens, and, especially, David Neumark for their comments on this paper. For their help in obtaining the data for this study, he also thanks the many American Express Financial Advisors he consulted.

Data and programs used in this study can be obtained on request to the author at 4610 University Avenue, Suite 700, Madison, WI 53705-2164. E-mail address: DANH@LRCA.COM.101.

Industrial and Labor Relations Review, Vol. 51, No. 1 (October 1997).
0019-7939/97/5101 $01.00

ming errors in the example above), but the workers share equally in the gain. Unlike with individual incentives, the workers are not pitted against one another, so any potential gains from cooperation should be realized.

Group incentives, however, are plagued by the free rider problem. Under individual incentives, if there are N workers in the firm, and one employee improves his or her productivity by $100, this will (or could) lead to a $100 reward. But if the gains are shared with the work group, this $100 gain in productivity only leads to a $100/N gain to the worker. As the size of the group grows, the individual incentive to expend more effort decreases. Furthermore, if it is assumed that the marginal cost of effort increases with effort and workers have identical cost-of-effort functions, the hardest workers are the least likely to respond to the incentive plan. That is, the workers who are expending the highest levels of effort prior to the implementation of the incentive plan require a larger reward before they will increase their effort.

Most previous studies of the effects of these incentive plans have examined firm-level data and have used value added as the dependent variable. Weitzman and Kruse (1990) provided a thorough survey of the literature on both profit sharing and gain sharing plans (which they considered to be fundamentally the same). They concluded that the "mean estimated effect of profit sharing on productivity for 'average' amounts of profit sharing ... is 7.4 percent, with a median estimate of 4.4 percent" (p. 138).

In contrast, in this paper I examine *individual* responses to the introduction of a gain sharing plan. The examination of individual-level data allows a determination of both the overall response to the incentive plan (as in firm-level studies) and the variation of the response to the incentive plan across workers of different types (for example, those who were productive prior to the implementation of the incentive plan versus those who were unproductive). I use data from two units within American Express Financial Advisors, a subsidiary of the American Express Corporation.

Previous Research

The only study comparable to this one appears to be Weiss (1987).[1] Weiss studied three plants within a large U.S. electronics manufacturer. At these plants, newly hired workers faced an individual incentive, but once they reached a performance threshold—usually occurring after three months—their pay was proportional to the output of their work group. Weiss estimated the change in individual output when the workers changed incentive schemes.

The study found that "almost all of the workers whose first month performance was above the median decreased their output between months 1 and 4,... [but] only half of the low performing workers decreased their output" (p. 140). In addition, "Among workers whose first month output was more than 10% below the median, 79% increased their output. Out of the 208 workers whose first month output was more than 10% above the median only one worker increased his output" (p. 140). In other words, the group incentive failed to motivate the majority of the workers, and workers' performance converged to a standard.

Weiss speculated that the most able workers were held back by both "trivial financial incentives" and "pressure from other workers" (p. 143). Brown (1990) summarized Weiss's work as "strong evidence that in large groups these incentive effects may be lost" (p. 174-S).

Data and Setting

The two units to be examined comprise the service representatives for the discount brokerage section of American Express Fi-

[1] A recent paper by Shearer (1996) used individual-level data from a firm using a group incentive plan. However, the goal of that paper was to estimate a tenure-productivity profile *controlling for* the incentive effects, and not to directly estimate the impact of the group incentive.

nancial Advisors. The units perform identical tasks, and are physically adjacent. One unit is responsible for clients in the Western half of the United States, and the other covers the remainder of the country. Each unit has its own supervisor, who reports to still another manager who oversees both units. The units are rated jointly to determine the size of their bonus. Although they could operate in isolation, in practice there is a great deal of cooperation between the units, so a joint rating is not inappropriate.

The workers are primarily responsible for data entry, bookkeeping, and receiving calls from clients.[2] There are two distinct phone lines to which a worker may be assigned: the service line and the trading line. New transactions (the actual buying and selling of securities) take place on the trading line. As this is a discount brokerage firm, it is not the worker's job to give advice or promote sales. For this reason, conversations on the trading line follow a standard process of information gathering (finding out how many shares to trade, and so on). The service line handles a wider variety of calls, ranging from simple questions about the status of a trade to questions (concerning trade errors, for example) that demand complex responses. All of the workers are qualified to work on the service line, but only some of them have obtained the licensing required to work on the trading line.

This study examines only performance on the service line. This is limiting in that we cannot determine whether any measured response in service line work comes at the expense of the other tasks.[3] As is noted below, however, the specific goals of the incentive plan focus on improving performance on this line, and in fact those goals are the only quantifiable ones. Therefore, an analysis of the impact of the plan using telephone performance data should be sufficient for determining whether the incentive plan succeeded.

The telephone performance data do not depend on individual reporting. In order to take calls off the queue, workers log onto a system, after which the number of calls answered, the amount of time spent on calls, the amount of time available to take calls (that is, the time the worker is logged onto the system), and the number of calls transferred are all recorded. The numbers are totaled for each worker on a weekly basis and sent to the supervisor. Only monthly data were available for use here. The average minutes spent per call in a month (abbreviated "minutes per call" hereafter) is used as a proxy for productivity. Workers who take fewer minutes per call can handle a higher volume of calls, and are therefore considered more productive. There are some potential problems with this interpretation, however, which will become apparent below.

The distribution of incoming calls across workers is random, so in principle the variety among types of calls should not affect the average minutes per call across workers. However, a worker can choose to transfer a problem to someone who is considered more able. Although any call that a worker receives from the queue is included in the data, a call received via transfer from another worker is not recorded. Since transferred calls are likely to be short, there may be a bias in minutes per call, associating less able workers with lower minutes per call. In the empirical analysis, we take this into account by adding the percentage of calls that a worker transfers to other workers as a control variable.

Workers differ substantially in their methods. Some prefer to call back clients instead of leaving them on hold, some provide clients with more thorough (and therefore lengthy) explanations, and so on. These differences make it difficult for supervisors to judge relative performance. If a worker is taking more time per call it could mean that better service is being

[2]The information regarding details of the incentive plan, work conditions, and responsibilities of the workers was obtained while I was employed with this firm. I was a temporary employee, however, and therefore not involved in the incentive program.

[3]One possible advantage of Weiss (1987) over this study is that the performance measure Weiss used may be a better measure of a worker's total output on the job.

given, that fewer problem calls are being transferred to other workers, or that he or she simply likes to chat with clients.

Given these problems with the data, one must be careful in interpreting what the variable *minutes per call* can say about worker performance. That worker A's average minutes per call exceed worker B's does not necessarily imply that worker A is less productive. It is true that more time per call means fewer calls can be answered, but if worker A deals with more complex problems than worker B, he or she could be more valuable.

Although minutes per call may not be as informative as one would like, the change in minutes per call for a given worker when the incentive plan is implemented does indicate whether effort was exerted to fulfill the established goals. That is, the primary goal of the plan was to reduce the amount of time a client must hold before speaking to a representative, and from an individual standpoint, the only way to achieve this is to spend less time on each call, thereby getting to the calls in the queue faster. Therefore, if workers take less time per call after the implementation of the incentive, they have responded positively to the incentive.

The details of the incentive plan are as follows. It was conceived in periodic meetings between managers, supervisors, and workers in late 1991, who discussed the most pressing needs (or shortcomings) of the department and the best means of addressing them. Following these meetings, specific goals were agreed upon, put in print, and distributed to all of the concerned parties.

The bonus amount was based on a unit rating given by upper management, and all members of the unit, including supervisors, were eligible for the bonus. Workers hired during 1992 were eligible for a prorated bonus. The awards were to be paid in early 1993, with the highest rating paying each worker $1,150, the second highest $800, and the third highest $600.[4] There were two lower ratings that did not merit bonus pay. All workers in the unit who met a minimum individual performance rating (given annually by the supervisors) were to receive the same bonus, based on the performance of the unit. The minimum individual performance rating requirements probably were not stringent enough to preclude the possibility of free riding, as the lowest ratings were the equivalent of being placed on probation and were rarely given. All of the workers in the sample were given performance ratings that exceeded the minimum.

As stated above, the plan focused on reducing the amount of time callers had to wait to receive service. This is measured at the unit level by the "ASA," or average speed of answer.[5] In June 1991, the ASA was about 79, meaning that on average, callers waited 79 seconds before speaking to anyone. The targets for the plan were as follows: a reduction to 35 for the highest rating, 55 for the second-highest, and 65 for the third-highest.[6]

The Model and Empirical Results

The dependent variable in all of the following specifications is the monthly average minutes spent per call by each worker. The independent variables control for changes in the workplace conditions such as supervisory policy and demand conditions. Data are not available to determine

[4]Since the reward schedule is specified in advance, the nature of the free rider problem in this firm is somewhat different from that described in the introduction. That is, any gains in individual productivity are not directly shared with the other workers, but contribute toward the team goal. Free riding is still an issue, however, since a worker who contributed little to the plan would receive the same reward as the other workers if the plan proved successful.

[5]Because the ASA is only available at the unit level, but the analysis is conducted at the individual level, minutes per call is used as the proxy for productivity.

[6]Other, more qualitative goals were mentioned. One example: "Actively develop a pro-active vs. reactive listening approach." Given the difficulty of measuring any progress toward such a goal, it is easy to see why managers would focus on changes in the ASA as the definition of success or failure of the plan.

Table 1. Means of the Data by Year.

Variable	*1991*	*1992*
Minutes per Call	4.86	4.22
Percent of Worker's Time Logged on the Phone, but Not on a Call	.22	.24
Percentage of the Unit's Volume Taken by the Worker	.09	.05
Percentage of Calls Transferred by the Worker	.26	.21
Average Number of Workers in the Unit	9.4	13
Average Number of Workers in the Sample (Two Units Combined)	18.7	15.8
Average Monthly Volume of Calls Received by the Unit	5,411	6,277
Male	.43	.43
High School	.10	.10
Associate's	.10	.10
Bachelor's	.80	.80
Tenure (Months)	11.9	18.9
Pre-Company Experience (Years)	7.2	7.2
Nominal Wage (Hourly)	12.41	12.92

if a change in supervisors occurred in the sample period, but information on supervisory policies does exist. That is, there are two actions that a supervisor may take to reduce a client's time on hold: hire more workers, or schedule more telephone time for the existing workers. In fact, as Table 1 shows, the number of workers did increase over the sample period, from about ten workers per unit in 1991 to thirteen workers per unit in 1992.[7]

A change in the allocation of a worker's time is indirectly measured by the percentage of a worker's time spent logged on the phone (that is, available to take a call) but not speaking to a client (abbreviated PERCTIME). Telephone time was scheduled for each day, and supervisors may have increased the amount of time scheduled in order to help meet the goal of reducing the ASA. PERCTIME increased between 1991 and 1992, a finding one would expect to observe if more time was allocated to the telephones.

The monthly volume of calls received by each unit controls for changes in behavior due solely to fluctuations in the amount of business received. Table 1 shows that the volume of business was, on average, increasing over the sample period, presenting the possibility that workers hurried through calls simply because more calls were coming into the queue.

An examination of the workers' characteristics in Table 1 reveals that there were slightly more women than men, the vast majority were college-educated, and workers had an average of about seven years of pre-company experience.[8] The average nominal wage was between $12 and $13 per hour.

The data are monthly, from June 1991 through November 1992. Although a total of 42 people were employed in the two units during this time period, only 21 of them are in the sample both before and after the implementation of the incentive plan.[9] The average number of workers per

[7]This count includes all of the workers in the units, including those not in the sample.

[8]Pre-company experience was estimated by age – tenure – education – 6, where a high school degree was equivalent to 12 years, an associate's 14, and a bachelor's 16.

[9]Because complete data on the entire work force were not made available, I could not include all of the workers in the study.

Table 2. Fixed Effects Estimates of the Effect of the Incentive Plan.

Variable	*(1)*	*(2)*	*(3)*	*(4)*
Incentive Plan	−.134**	−.184**	−.174**	−.100*
	(.031)	(.043)	(.043)	(.061)
Serial Correlation Correction	No	No	Yes	Yes
Trend	—	—	—	−.019**
				(.009)
Percent of Worker's Time Logged on the Phone, but Not on a Call	—	−.620**	−.547**	−.467**
		(.113)	(.117)	(.120)
Percentage of the Unit's Volume Taken by the Worker	—	−.979**	−.925**	−1.085**
		(.361)	(.366)	(.370)
Percentage of Calls Transferred by the Worker	—	−.613**	−.726**	−.781**
		(.290)	(.300)	(.304)
Number of Workers in the Unit	—	−.007	−.0004	.015
		(.010)	(.012)	(.014)
Log(Volume of Calls Received by the Unit)	—	.115	.114	.038
		(.083)	(.077)	(.085)
N	290	290	266	266

Notes: Standard errors are in parentheses. Log(minutes per call) is the dependent variable; therefore, lower numbers mean higher productivity. All models include a constant. The incentive plan variable is a dummy variable equaling one for months in 1992. Observations are lost in the correction for serial correlation because of breaks in the data series for three of the workers.

*Statistically significant at the .10 level; **at the .05 level (two-tailed tests).

unit in Table 1 refers to the *actual* number of workers per unit. This variable, and not the number of workers in the sample, is used in the analysis because it is the actual number of workers that could affect the behavior of the workers. The average number of workers *in the sample* is also shown in Table 1, revealing that on average there are fewer workers in the sample in 1992 than in 1991.

The basic form of the model that is estimated is:

$$(1) \qquad \ln(\text{minutes per call})_{it} = \Sigma_i \alpha_i I_i + \beta_1(\text{incentive plan})_t + \beta_2(\text{workplace controls})_{it} + \varepsilon_{it},$$

where i indexes workers, t indexes months, and I_i are dummy variables for individuals. The model is run both with and without the workplace controls. The "incentive plan" variable is a dummy variable equaling zero for months in 1991 and one for months in 1992. This specification of the incentive plan may be considered a shortcoming of this study because other changes affecting the productivity of workers may have occurred during this time, and if they did, the dummy for the incentive plan may be picking up these other effects.

The possibility that the results are contaminated in this way is minimized for several reasons. First, due to a reorganization, the data do not begin until June 1991. It was at this time that the workers split into two units handling accounts from different regions of the country. No other reorganization occurred during the sample period. Second, the analysis controls for the monthly volume of calls to account for any demand fluctuations. Third, variables controlling for supervisory behavior are included. Specifically, variables are included to reflect the fact that more workers were hired and that the time within the work day was reallocated.[10]

[10]The supervisors could also increase the level of their monitoring, but such a change cannot be measured. A possible problem with the analysis is that the supervisory policies that can be included (hiring and reallocating time) may lead to higher minutes per call for existing workers, while the excluded action (increased monitoring) should reduce minutes per

Table 3. Fixed Effects Estimates Testing for Sample Attrition Bias.

	Sample Cut-off Date				
Description	*February 1992*	*March 1992*	*May 1992*	*August 1992*	*November 1992*
Incentive Plan	−.109*	−.106*	−.149**	−.151**	−.184**
	(.059)	(.057)	(.054)	(.048)	(.043)
Number of Workers in the Sample at the Cut-Off Date	21	16	16	14	11
N	169	185	216	260	290

Notes: Standard errors are in parentheses. Log(minutes per call) is the dependent variable; therefore, lower numbers mean higher productivity. All models include a constant. The incentive plan variable is a dummy variable equaling one for months in 1992. The specification used is the same as the one used in column 2 of Table 2.

*Statistically significant at the .10 level; **at the .05 level (two-tailed tests).

The first set of regression results is reported in Table 2. Fixed effects estimation is used throughout. An F-test of the joint significance of the individual dummies produces an F-statistic of 11.22, rejecting the null hypothesis that the pooled model is efficient.

Column 1 shows the effect of the incentive plan with no control variables included. On average, minutes spent per call went down by about 13% in 1992. Column 2 adds the workplace controls, increasing the estimated effect of the incentive plan. The estimates of the workplace controls are generally highly significant, and consistent across models.

A potential problem arises from the fact that the panel of workers is unbalanced, raising the possibility that attrition from the panel may be affecting the results. That is, one may falsely observe that the incentive plan was successful if less productive workers are in the sample prior to the beginning of the plan, but leave soon after implementation. To test for this possibility, I divided the sample into two groups: those who performed better than and worse than the median level in 1991.[11] A worker's presence in the panel is represented by two variables, the number of months present in the data in 1991 ("LENGTH91") and the number of months present in the data in 1992 ("LENGTH92").[12] A test for the equivalence of means between the two groups is done for these variables. In neither case does the test reject the hypothesis that the means for the two groups are equal.[13] Since neither group is greatly over-represented either before or after the incentive plan is introduced, one cannot make the claim that an improvement may be observed only because inferior workers were weeded out of the sample.

call. This may bias the results toward finding a large independent response from workers to the incentive plan. To explore that possibility, I duplicated the entire analysis excluding all supervisory behavior (dropping PERCTIME and number of workers in the unit from the regressions). The results from these regressions show a stronger estimated response in most specifications, with decreases in the estimates (in magnitude and significance) occurring only when a time trend is included.

[11]This division is based on the fact that the estimated response is driven primarily by the workers who performed worse than the median in 1991.

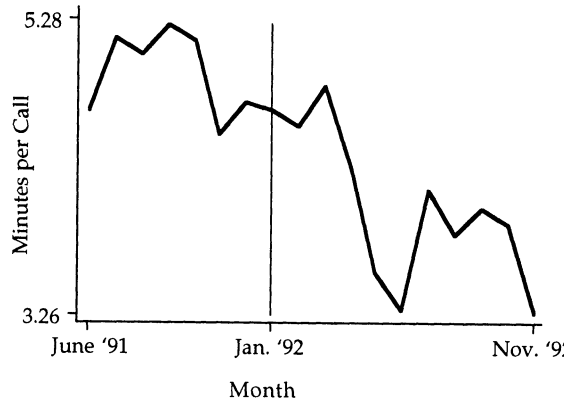
[12]There are 18 months of data, beginning in June 1991. Thus, for example, a worker who was in the data from November 1991 through March 1992 would have LENGTH91 = 2, LENGTH92 = 3.

[13]The means are as follows (workers better than median, workers worse than median): LENGTH91 (5.9,6.3) and LENGTH92 (7.5,7.7).

A second test for sample attrition bias is conducted by observing how the results change as the sample period is shortened. The results (using the specification in column 2 of Table 2) are shown in Table 3. Notice that regardless of when the sample is cut off, the qualitative finding of a strong response to the incentive plan remains. In addition, the way the estimated effect of the incentive plan evolves in relation to the sample size shows that sample attrition bias is not driving the results. In February 1992, the sample size is at its highest point, with all 21 workers in the sample. If the data are cut off at this point, the estimated effect of the incentive plan is an improvement of about 11%. Five workers leave the sample from February to March, the largest drop between any two months. Yet, the estimated effect of the incentive plan using data through March is almost identical to the February estimate. Furthermore, no additional workers are lost through May 1992, yet the estimated effect of the incentive plan increases to about 15%. Given these facts, it does not appear that the estimated improvements in minutes per call are due to attrition from the sample.

Columns 4 and 5 of Table 2 refine the previous estimates by correcting for serial correlation and including a time trend. Quasi first-differencing is performed in order to correct for serial correlation using a separate value of ρ for each worker.[14] As shown in column 3, the finding that the incentive plan was a success withstands the correction for serial correlation. Furthermore, while the addition of a linear time trend reduces both the magnitude and significance of the estimated effect of the incentive plan, the estimate is still sizable and significant at the 10% level.

Figure 1. Average Minutes per Call by Month.



It seems that the finding of a significant response to the incentive plan is fairly robust. Given that the variable representing the incentive plan is merely an indicator equaling one for months in 1992, this result shows that the average minutes per call dropped significantly across years, but gives no indication about the timing of the impact. Figure 1, which graphs the pooled average minutes per call across months, shows clearly that it took several months for a large effect to occur. I do not know of any excluded variable that would cause a large change from March through May.[15] The volume of business was increasing, as was the number of employees, but those changes have been controlled for in the regressions. One possible explanation is that free riding occurred in the period shortly after the implementation of the incentive plan.

[14]The quasi first-differencing is performed as follows: ρ is estimated as the coefficient on the residuals from month $t-1$ regressed on the residuals from month t (separately for each worker). The data are then transformed by $x_t - \rho(x_{t-1})$. The first observation for each individual and any observations following gaps in the data series within individuals are dropped in this process.

[15]Changing the specification of the incentive plan dummy variable to begin one to two months before and after the actual implementation produces the following coefficients (using the specification from column 2 of Table 2). November 1991: –.135; December 1991: –.139; January 1992 (actual start date): –.184; February 1992: –.158; March 1992: –.160. Note that the largest estimated effect of the incentive plan occurs at the actual start date, as one would expect if the estimates are picking up the effect of the incentive plan, and not some other event.

Table 4. Fixed Effects Estimates of the Separate Samples.

	No Serial Correlation Correction		*Corrected for Serial Correlation*	
Description	*Initially Better Workers*	*Initially Worse Workers*	*Initially Better Workers*	*Initially Worse Workers*
Incentive Plan	.007	−.339**	−.014	−.173*
	(.051)	(.065)	(.067)	(.095)
Trend	—	—	.001	−.033**
			(.011)	(.014)
Percent of Worker's Time Logged on the Phone, but Not on a Call	−.358**	−.889**	−.309**	−.818**
	(.121)	(.187)	(.113)	(.224)
Percentage of the Unit's Volume Taken by the Worker	−.332	−2.060**	−.090	−2.103**
	(.390)	(.625)	(.385)	(.633)
Percentage of Calls Transferred by the Worker	−.645*	−.419	−.884**	−.315
	(.339)	(.431)	(.345)	(.488)
Number of Workers in the Unit	−.009	−.019	.002	.015
	(.012)	(.016)	(.016)	(.023)
Log(Volume of Calls Received by the Unit)	.048	.221*	.031	.140
	(.100)	(.125)	(.107)	(.135)
N	134	140	122	129

Notes: Standard errors are in parentheses. Log(minutes per call) is the dependent variable; therefore, lower numbers mean higher productivity. All models include a constant. The incentive plan variable is a dummy variable equaling one for months in 1992. Observations are lost in the correction for serial correlation because of breaks in the data series for three of the workers.

*Statistically significant at the .10 level; **at the .05 level (two-tailed tests).

It may be rational for the workers to attempt to free ride early on, as an *immediate* response is not required for the incentive plan to be considered a success. This could be taken as evidence that the free rider problem may be overcome in a repeated game setting, as was theorized in Weitzman and Kruse (1990).

The next issue to explore is whether the workers who were initially more productive responded as strongly to the incentive plan as did the other workers. For this analysis, the average minutes per call is calculated for each worker and for the total unit for 1991. The sample is then divided into those whose initial minutes per call were better and worse than the median.[16] Table 4 shows the estimates for these two groups. The first two columns can be compared to column 2 of Table 2, while the second two columns employ the specification used in column 4 of Table 2. These estimates show that the initially slow workers responded strongly to the incentive plan, but the initially fast workers showed no response at all.

Figure 2 shows the change in the productivity of individual workers graphically. The horizontal axis measures the average minutes per call for each worker in 1991, while the vertical axis measures the change in minutes per call from 1991 to 1992, calculated so that a positive number indicates an improvement (reduction in minutes per call). The vertical line is drawn at the 1991 median level of minutes per call. Of the ten workers who were better than the median in 1991 (to the left of the vertical line), six showed a decline in performance while only three improved.[17] However, all of the

[16]The worker at the median is dropped from the sample.

[17]Only one of these differences is significantly different from 0, however.

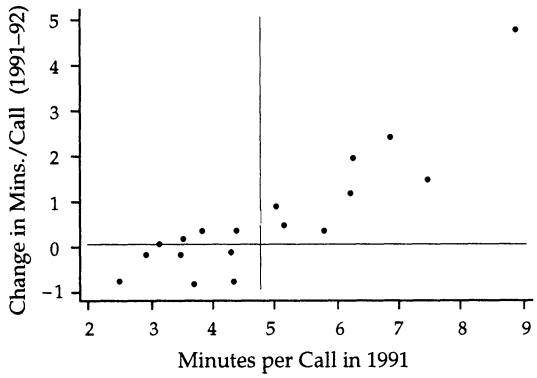
ten workers who performed worse than the median in 1991 improved their performance.

There are several possible explanations for the fact that the initially faster workers did not improve in 1992. First, this finding could support Weiss's (1987) conclusion that group incentives do not motivate the "best" employees. Second, it could be that there is a minimum amount of time that is required to provide service, so that those who were already performing well had very little room to improve.

Third, given that the calls are initially randomly distributed, it is possible that this effect merely reflects a regression to the mean. That is, the workers who were slow in the early months may have been as productive as other workers, but "unlucky" in that they happened to draw long calls off the queue (although such a systematic pattern is unlikely given the large number of calls each worker received in a month). To see if such a pattern might be affecting the results, I created a dummy variable for those who were slower than the median worker from July through September of 1991, then ran a regression using the previous specification, but including the newly created dummy, for October–December 1991. The use of only 1991 data ensures that the incentive plan is not affecting the test. This regression produces a coefficient of .529 (with a standard error of .093) on the dummy variable for the initially slow workers, showing that workers who began the sample by taking more time on calls continued to be slow in the following months. Given this finding, the results do not seem to be affected by regression to the mean.

Finally, an Akerlof-type gift exchange model might explain why the better workers seemed to ignore the incentive plan. If they felt that the slower workers were made to look bad by the disparity in performance, they may have attempted to lessen the difference by waiting for the slower workers to improve before attempting such an improvement themselves (Akerlof 1986). According to Akerlof, this represents a gift exchange between workers that occurs because "in working together, workers acquire sentiment for each other" (Akerlof 1986:74). The situation here differs somewhat from Akerlof's description, however. The workers that he describes are performing in excess of the established standard as a gift to the firm. In exchange, the firm must not raise the standard, ensuring that the marginal workers meet the required performance standard.

Figure 2. Changes in Individuals' Minutes per Call.



In contrast, the "better" workers studied here are theorized to be refraining from raising the standard, or helping the less efficient workers by handling the more difficult calls (via transfer, which would not be recorded in the data). In exchange, the "worse" workers are responsible for increasing their level of performance enough to meet the goals of the incentive plan. The firm receives the gift of faster service, and in exchange must pay all of the workers for any improvements. Frequent feedback regarding the progress made toward lowering the ASA (recall that telephone performance data were available each week) makes this exchange among workers possible. If, after a few months, the goal appeared to be unattainable under the arrangement, the faster workers could simply speed up and achieve the goal. Therefore, the gift that the faster workers are attempting to give may not jeopardize their chances for the bonus pay.

The final issue to be addressed here is

whether the observed change in minutes per call was purely cosmetic, or a real improvement on the part of the workers. In principle, minutes per call could easily be manipulated by answering the phone, taking a message, and then returning the call. Only the short initial call would be recorded as data, while the total time spent with the client may have increased. However, it is unlikely that this occurred, since the managers are not directly concerned with a change in minutes per call, but with the average speed of answer (ASA), which is not as easy for a worker to manipulate without a real improvement.

There is one way that ASA can be manipulated, but managers were not only aware of this possibility, but actually tried to use it to their advantage the previous year. As one might expect, this job faces a peak-load problem. During peaks, wait times for callers may increase dramatically. In the spring of 1991, a "release valve" was created in an attempt to solve this problem. One person would log onto the telephone queue and only take brief messages for the calls answered. The messages were then distributed to the workers so that the calls could be returned when business slowed. Note that the use of this method may result in a reduction of the ASA, even though the clients do not receive service any sooner (nor, possibly, any more efficiently). The release valve was considered to have benefits in spite of this because it alleviated the frustration clients felt from remaining on hold for an extended period of time.

Although the theory behind the release valve was sound, it ultimately failed because business rarely slowed enough to allow time to return the calls. Because of the ensuing complaints received by the managers from angry clients whose calls were not returned, the method was abandoned within months of its inception. Its existence, however, signals that the managers knew how the ASA could be manipulated, and that this method would produce complaints that revealed its use. The fact that, despite this knowledge, they chose to award bonuses suggests that the incentive plan caused a real improvement in service.

Comparison to Other Studies

Given that I have found that a group incentive can increase the average level of productivity but Weiss (1987) did not, it is instructive to compare the two environments that are studied. First, the sizes of the work groups studied are quite different. A total of 42 employees worked in the two units studied here over the sample period, versus an average work group of 126 in Weiss's paper. Since the free rider problem should worsen as work groups increase in size, this may be an important reason for the difference in findings.

There is, however, some evidence suggesting that size may not be negatively correlated with the success of an incentive plan. For example, Schuster (1984) used longitudinal plant-level data to examine the effect of a Scanlon Plan. This variant of gain sharing uses committees of workers to evaluate suggestions from other employees, and a bonus formula to share gains in productivity among the workers. In spite of the fact that there were 890 non-supervisory production employees in a plant,[18] Schuster found that the plan had a significant impact on productivity. He gave much of the credit for the plan's success to employee participation in decision-making. In addition, in his survey of the literature, Schuster concluded that "firm size correlated positively with rated Plan success and Plan retention" (Schuster 1984:25). This may be evidence that the specific form of the incentive plan, and not the size of the work groups, is what led to the difference in the measured success between this study and Weiss (1987).

Second, the incentive plan examined here differs in form from the one in Weiss (1987) in several ways. In the plan studied here, specific goals were set, so that the success or failure of the incentive plan could be easily determined. In Weiss (1987), the group incentive is not a goal-based incentive, but an everyday compensation for-

[18]The size of the work groups is not reported.

mula, in which an individual is placed in a group, and each worker receives pay in proportion to the output of his or her group. In addition, the firm studied in this paper used committees consisting of workers and managers to set the goals. These differences, which form a common theme in the industrial relations literature on group incentives, could be important.

Strauss (1990) listed some of the reasons that goal-setting and employee participation may raise productivity. They include the following: "Participation may result in better decisions"; "Participation may improve communication and cooperation"; and "Participation frequently results in the setting of goals. There is considerable evidence that goal setting is an effective motivational technique" (Strauss 1990:5). Schuster's (1984) finding that goal-setting and employee participation resulted in an increase in productivity provides empirical support for these assertions.

Summary and Conclusion

I have used individual level data from the discount brokerage section of a large firm to examine the impact of a group incentive plan. The results show that there was a significant unit level response, about a 17% improvement from the pre-plan average. Given the imperfection of the proxy for productivity and the fact that only one dimension of performance is examined, there is some question as to whether this can strictly be thought of as an increase in productivity. (Nor is it easy to determine whether the incentive plan was, strictly speaking, a financial success for the firm. While the payouts to workers can be calculated, it is not possible to place a monetary value on the change in worker behavior that occurred.) There can be little doubt, however, that at least a change in behavior occurred.

Although the group incentive did increase the average performance for the units, those workers who were initially the best may not have been motivated by the plan. Some of these workers did not merely show less improvement than those who were initially worse, but actually declined in performance after the introduction of the incentive. Here too, the limitation of the dependent variable makes the conclusion tenuous. That is, if those who were initially taking very little time per call were subsequently asked to handle more time-consuming problems or provide better service, they may have improved in a way that the data cannot measure. It is surely worthy of note, however, that the managers who evaluated the plan deemed it successful enough to merit the second-highest rating, giving each worker $800.[19]

The finding that performance converges to a standard was also observed in Weiss (1987). Given the many differences between the environment studied by Weiss and the environment studied here, this is a surprising similarity, and may indicate a pattern in workers' responses to group incentives.

However, Weiss (1987) can also be interpreted as showing that group incentives may be ineffective when work groups are large, whereas this study found an improvement in mean performance; and I have presented evidence from other studies suggesting that the design and implementation of the incentive plan, and not the difference in the size of the work groups, led to this disparity.

This paper therefore presents the possibility that a well-designed group incentive can increase the mean level of productivity in a variety of settings. An important caveat to this tentative conclusion is that the initially most productive workers may be largely unaffected (and possibly adversely affected) by the incentive.

[19]Recall that in order for employees to be given this rating, the incentive plan called for a reduction in the ASA from 79 seconds to 55, a 41% reduction. It may seem that the 17% reduction observed in individual minutes per call could not account for this, but the two statistics are not comparable in this way. Remember that ASA is a unit-level measure, and would also be affected by such things as hiring more workers and the allocation of more of each employee's day to the telephones.

REFERENCES

Akerlof, George A. 1986. "Labor Contracts as Partial Gift Exchange." In George A. Akerlof and Janet L. Yellen, eds., *Efficiency Wage Models of the Labor Market.* Cambridge: Cambridge University Press, pp. 66–92.

Brown, Charles. 1990. "Firms' Choice of Method of Pay." *Industrial and Labor Relations Review,* Vol. 43, pp. 165S–182S.

Greene, William H. 1990. *Econometric Analysis.* New York: Macmillan, pp. 440–41.

Schuster, Michael. 1984. "The Scanlon Plan: A Longitudinal Analysis." *Journal of Applied Behavioral Science,* Vol. 20, pp. 23–38.

Shearer, Bruce. 1996. "Piece-Rates, Principal-Agent Models, and Productivity Profiles: Parametric and Semi-Parametric Evidence from Payroll Records. *Journal of Human Resources,* Vol. 31, pp. 275–303.

Strauss, George. 1990. "Participatory and Gain-Sharing Systems: History and Hope." In Myron J. Roomkin, ed., *Profit Sharing and Gain Sharing.* Metuchen, N.J.: IMLR Press/Rutgers University, pp. 1–45.

Weiss, Andrew. 1987. "Incentives and Worker Behavior: Some Evidence." In Haig R. Nalbantian, ed., *Incentives, Cooperation, and Risk Sharing.* Totowa, N.J.: Rowman & Littlefield, pp. 137–50.

Weitzman, Martin L., and Douglas L. Kruse. 1990. "Profit Sharing and Productivity." In Alan S. Blinder, ed., *Paying for Productivity.* Washington, D.C.: Brookings Institution, pp. 95–140.