# A Study of Performance Measurement System Implementation in a High Volume Production Line

Eric O. Olsen<sup>1)</sup>, Honggeng Zhou<sup>1)</sup>, Yoke-Eng Ng<sup>2)</sup>, Chow Chewn Chong<sup>2)</sup>, <u>Pean Padunchwit</u><sup>2)</sup>,

<sup>1)</sup> The Ohio State University - Fisher College of Business (olsen.54@osu.edu)
 <sup>2)</sup> Hewlett-Packard Singapore - Inkjet Supplies Manufacturing (pean\_lim@hp.com)

#### Abstract

Most managers have heard the platitude "what gets measured gets done", but the fact is that there is little empirical research on what constitutes an effective performance measurement (PM) system. Internal performance measures provide critical infrastructure that allows an operation to 1) document historical performance, 2) indicate current position with respect to operational strategy, 3) predict future performance, and 4) motivate and control behavior. PM system effectiveness can be described as a system's ability to address all of these functions.

In order to develop grounded theory in PM system design and analysis, a single longitudinal case study was undertaken on a subassembly process in Hewlett Packard's high volume Inkjet pen cartridge production line in Singapore. A new PM system was introduced to "motivate operators to higher performance, provide for fairness and consistency in salary related performance evaluation, and encourage ownership in the operation". Measures emphasized workcentre yield, output, housekeeping, teamwork, suggestion system participation, quality workmanship, safety, and operator dependability (attendance) at the individual, team, and production line levels. Each operator was provided a monthly evaluation of his or her individual performance based on a point allocation system.

Twenty-seven months of pre and post PM system implementation yield and operator productivity data were compared. Overall, production yield showed a significant increase in both level and rate of improvement. The level of operator productivity also improved significantly over the period studied. Individual performance measure effectiveness with respect to causality, improvement, and control were analyzed. Conclusions are drawn on PM system effectiveness from this analysis and feedback from line management. Recommendations are made for a subsequent round of PM system improvements and future research.

# 1. Introduction

Internal performance measurement systems play a critical role in the implementation of strategy. However, little empirical research exists to inform practitioners exactly what constitutes an effective performance measurement (PM) system [1]. There is a large body of literature formed from consulting and practitioner recommendations and anecdotal evidence. Kaplan and Norton's [2-5] extensive work on using "balanced scorecards" to offset the pervasive reliance on strictly financial measures of performance is perhaps the leading concept in this stream. Other examples are referenced [6-9]. All these authors put forward a laundry list of attributes that they feel constitute an effective PM system. This paper seeks to build theory and describe a framework for analyzing the effectiveness of a PM system based on three important attributes, namely causality, improvement, and control.

This paper is organized as follows. First, the importance of the three attributes of PM effectiveness is supported from the literature. Next, a framework for analyzing the occurrence and extent of these attributes in real PM systems is proposed. This framework is then illustrated using 37 months of data including a newly implemented PM system in the Intermediate Assembly line of Hewlett Packard's high volume Inkjet pen cartridge production in Singapore. Finally, conclusions are offered along with recommendations for future research.

#### 2. Literature Review

Internal performance measures provide critical infrastructure that allows an operation to 1) document historical performance, 2) indicate current position with respect to operational strategy, 3) predict future performance, and 4) motivate and control behavior. PM system effectiveness can be described as a system's ability to address all of these functions.

# 2.1 Causality

The attribute of causality speaks to the PM system's relationship to operational objectives. Strategic alignment and a "tops down" perspective is key. Performance measures must be driven from the top, directly linked to the organization's strategy and vision" [10]. Measures should have a direct relationship to manufacturing strategy. Performance measures must directly measure the success or failure of each manufacturing strategy [6].

Performance measures need to have content validity. "Select performance measures that your business users value and use for guiding and supporting decisions" [9]. Blenkinsop and Burns [11] point out that PM's that if there is no direct connection between operational subunits and the performance of the company as a whole will cause the measure to be perceived as being irrelevant. PM's need to differentiate between special and common causes. Taking action on common cause variation as if it is a special cause can be detrimental to performance [12]. In other words, a PM system should also demonstrate predictive validity.

Causality is the degree to which the PM focuses on the production process and direct causes of performance as opposed to secondary effects. For example, scrap rate would be a direct cause of higher material and labor costs. Causality is closely associated with understanding the process of production [7, 12-13]. "Don't measure A and hope for B: An improved measure in one area doesn't always mean improvements in a seemingly related area. For example, measuring the time service representative spends on the phone with a customer doesn't necessarily serve as measure of improved customer service" [14]. Focusing on "process drivers" is a way to conceptualize causality [3, 12].

#### 2.2 Improvement

The improvement attribute describes a PM system's ability to capture trends in performance. Hayes, Wheelwright, and Clark [15] describe the purpose of performance measurement as describing a direction and rate of improvement over time. The PM system plays a key role in promoting continuous improvement. The PM system should to promote learning both at the individual and organizational levels [6, 16-20]. Plotting trends over time allows operators and managers to associate specific experiences [17]. A measurement system should focus on continuous improvement, rather than just compliance and control: While standards and requirements are clearly important, a system should also encourage measures and analyses that facilitate continuous improvement [14]. The PM system should "foster, rather than just monitor improvement" [6, 21].

#### 2.3 Control

A PM system must demonstrate the control over the targeted PM. Statistical process control (SPC) is a wellunderstood premise in controlling a process [16, 22, 23]. Variation in range and average sample values are maintained between statistically derived limits. As a process becomes more capable of meeting specifications, or in control, limits are reduced. The idea of control is related to the concept of ownership [9, 11, 14, 16, 24]. A work group should only be accountable for measures over which they have control. Because accountability and control are linked at the work group level, it is expected that the level and method of control will vary between workcentres. This should be reflected in the PM system design [6].

# 3. Analysis Framework

With the three PM system attributes of causality, improvement, and control, established as the criteria for an effectively designed PM system, it remains to operationalize a systematic method for assessing effectiveness. First, statistical models are developed for dependent variables that reflect the strategic objectives of the operation. External assessment data are best, but internal data may be used where the criteria for strategic alignment are met and there is a hierarchical relationship to the PM's. In the case of Hewlett Packard described below, cumulative yield and productivity data from the company's ERP (enterprise resource planning system) was used as representative of the operations strategic goals to excel in quality and cost. General linear models were developed for both dependent variables regressed against time and PM system implementation status. Once a statistically significant model is established, it is possible to assess whether the dependent variables themselves demonstrate the attributes of improvement and control. Improvement was assessed by a statistical comparison of the rate of improvement before and after implementation was used to assess control.

Once the overall impact of the PM system implementation on the strategic variables is described, it is then possible to examine the individual PM's to assess the role they play. Individual correlations are assessed for each PM and each dependent variable. Although correlation does not imply causality [25], it is proper to assume that the lack of significant correlation denotes the absence of causality. Therefore, correlation can be used to eliminate measures that lack the attribute of causality.

The attribute of improvement is assessed by the PM's correlation to time. If the correlation is significant, and in the appropriate direction with respect to the dependent variables, then it demonstrates the attribute of improvement. A trend exists beyond random fluctuation with time.

The attribute of control is demonstrated by a significant reduction in the variability of the measure. In cases where a significant trend exists, the mean absolute deviation (MAD) between the trend and the actual data was compared for the early months of PM system implementation versus the later months. If a specific target exists for the PM, as was the case for production output, the deviation from target was used. The mean value of the data was used in cases where

neither a trend nor a target existed. A convention of always making the early months the smaller when an odd number of months were available was used to avoid bias in sample selection. A single tail t-test assuming unequal variances was used to test for a significant decrease in a PM's variability. Rejecting the hypothesis of no difference implies improved control.

A couple limitations of this study resulting from the availability of data should be noted. The first is the fact that data was not available for most of the new PM's prior to implementation. It would have been nice to assess all three effectiveness attributes prior to PM system implementation. Second, even though 8-13 months of PM stem data were available, this is not a large sample size for the statistical analyses performed. Future research will incorporate nonparametric statistical techniques commensurate with smaller sample sizes. Finally, the fact that not all data was available for all the months included in the study is an artifact of working with real company data. People leave and standalone databases are subject to loss. As appropriate, comments will be made with respect to data availability in the analysis section.

# 4. PM System Design - A Case Study

A case study of Hewlett Packard's InkJet Intermediate Assembly (IA) operation in Singapore is used to illustrate the proposed PM system effectiveness analysis method. In July of 2000, IA management implemented a new PM system they designated as "Top Grading" (TG). The new PM system was introduced to "motivate operators to higher performance, provide for fairness and consistency in salary related performance evaluation, and encourage ownership in the operation". Simultaneously, IA's operations strategy places a high emphasis on quality and cost. As the middle operation in a 3-stage internal supply chain between integrated circuit manufacturing and final assembly, IA plays a critical role in assuring that cost and quality targets are met. The entire Singapore operation is tracked via an ERP (SAP) system along with its sister operations in the USA, Ireland, and Puerto Rico. Upper management views yield and productivity as crucial to achieving worldwide quality and cost goals. Yield is cumulative and is measured as ratio of total good units out of the production line to total units started expressed as a percentage. This value is supplied by the ERP system. Productivity data is also supplied by the ERP system. Productivity is measured as the ratio of production moves required for all products completed to the average number of full time equivalent operators including overtime. These two values, cumulative (cum) yield and productivity, are used as the dependent, strategic variables in the analysis.

The TG system consisted of a point-scoring scheme that awarded points to individual operators based on measure performance at the production line, workcentre, and individual performance levels. The measures and scoring criteria are shown in an actual scoring sheet in Figure 1. A more detailed description of the measures is included in Table 1. The PM system was designed with the intention of being effective, yet easy to understand and administer.

In October of 2001, an assessment of the effectiveness of TG metrics was initiated. The yield goal exceeded expectations while other measures seemed to lag. The study described in this paper was used to assisted line management in evaluating the PM system at the shop floor level. A second phase PM system is being designed to improve on effectiveness and increase the satisfaction of the production team. In general, management perceives that the implementation of this system has increased the satisfaction level of the operators with the overall fairness of their performance evaluation. Morale has increased while exceeding quality goals and continuing to increase productivity.

It is the authors' experience that PM systems are rarely initiated in isolation. This is supported by the "process view" of strategy where internal performance measures provide valuable feedback on the implementation of strategic initiatives [1]. Internal performance measurement is also integral to the daily support of continuous improvement activates in a world-class manufacturing environment. This is certainly the case in IA. Most significantly, during the period studied, IA implemented a visual control information system for metering their production output and controlling work in process inventory between workcentres [26]. They also emphasized systematic operator cross-training, empowered teams, reward and recognition. All of these efforts worked in concert with the TG system. Although we focus in this paper on the specific impact of the PM system and its characteristics, it should be recognized that ultimately it is the actions motivated by the PM system that cause the changes in the strategic variable results.

# 5. Analysis:

#### 5.1 Strategic (Dependent) Variables

Cumulative yield and productivity were analyzed first to assess the overall impact of TG implementation on strategic performance. Both variables were modeled using general linear regression and tested for changes rate of improvement and for the level of variability before and after implementation.

Intermediate Assembly

Month \_\_\_\_\_ Operator Name \_\_\_\_\_ Supervisor Name \_\_\_\_\_

Monthly Work Performance							Assessment		
	Workcentre								
Criteria	Category	Measures	Grade			Target	Actual	Score	
Quality	Workcentre	Yield	Above Target 25pts	On Target 15pts	Below Target Opt				
		Above Stretch Goal (Bonus Pts)	Yes 10pts	No Opt					
Productivity	Whole Team	Output	Above Plan 15 pts	On Plan 15pts	Below Plan 0pts				
Material handling and Generic Housekeeping	Workcentre	Station cleanliness and proper storage of good, on- hold, and scrap materials.	Refer						
Team work	Individual	Able to develop good relationship with work group and cooperative.	Refe						
ESS (In Group)	Individual	Idea contribution (no limit)	Receive Awards 15pts	Participate 10pts	No response 0pt				
Workmanship	Individual	quality Rejects caused by workmanship problem	Deduction of 25pts						
Safety	Individual	Violate safety rules or regulations	Deduction of 25pts						
Dependability	Individual	No show for work	Deduction of 10pts						
					Total				

# Fig. 1 Top Grading performance measurement system scoring sheet

Variable	Units	Description					
Dependent	<u>.</u>						
Cumulative Yield	%	Ratio of total good units out of the production line to total units started expressed as a percentage. This value is supplied by SAP.					
Productivity	Moves/adjusted production FTE	Ratio of production moves required for all products completed to the average number of full time equivalent operators including overtime. This value is supplied by SAP.					
Independent	<u>.</u>						
Workcentre Yield (to target)	%	Ratio of total good units out of the workcentre to total units started into the workcentre expressed as a percentage. This value is reported by SAP. Points are score against the target yield per workcentre.					
Workcentre Yield (to stretch goal))	%	Ratio of total good units out of the workcentre to total units started into the workcentre expressed as a percentage. This value is reported by SAP. Points are score against the stretch yield goal per workcentre.					
Output	units	Total good units out of the entire production line.					
Housekeeping	points	A list of 15 housekeeping criteria is score as 1 or zero by a third party audit per workcentre for each of 4 shifts. The average of the 4 scores is designated as the workcentre's monthly score. Criteria range from smocks properly hung in the clear room to work in process inventory being properly labeled.					
Teamwork	points	A monthly peer review scored each operator on 5 teamwork criteria. The score was the total average input for each operator. Criteria range from willingness to help others to peer respect.					
Employee Suggestions	number	The number of written submissions to the employee suggestion system is tracked. The point value is increased if the submission receives a gold, silver, or bronze award.					
Workmanship	number of defective units	Any reject due to human error is counted as a point reduction. The supervisor tracks this.					
Safety Incidents	number of incidents	Any reportable or non-reportable safety incident. The supervisor tracks this.					
Dependability	number of incidents	Any deviation for attendance policy such as lateness, urgent leave, etc.					
Total Score	points	Total of points allotted for performance to each of the previous 9 measures.					

 Table 1 Detailed description of performance measures

(1) Cumulative Yield

Eighteen months of data were available before TG implementation and 19 months after implementation. An overall graph of the data is included in Figure 2. It is obvious that the behavior of the yield data is different before and after the implementation of the new PM system. TG was implemented in month 19.



Fig. 2 Graph of cumulative yield data pre and post Top Grading implementation

A regression model was developed to test characteristics of the data. The form of the model is:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$$

where

Y = predicted value of yield  $x_1 = \text{sequential month}$  $x_2 = 0 \text{ pre implementation or 1 post implement of TG.}$ 

The overall model is significant at the p=.000 level with an adjusted R<sup>2</sup> of 82.5%. The coefficient matrix is as follows (\*\* = significant at the .01 level):  $b_0 = .782^{**}$ ,  $b_1 = -.000144$ ,  $b_2 = -.00905$ ,  $b_3 = .000770^{**}$ 

The main effect of TG is not significant  $\phi_2 = -.00905$ ). This essentially means that there is no discernable difference in the overall level of cum yield with implementation. The high variability of the pre-implementation data makes achieving a significant result for a main effect less likely. However, a strong positive result is achieved for the interaction term ( $b_3 = .000770$ ). This validates the obvious difference in improvement trends before and after implementation. Prior to TG the trend was negative, but not statistically significant from zero ( $b_1 = .000144$ ). After implementation, the trend is statistically significant and positive. Every month since implementation, the yield has improved an average of .082%. The adjusted R<sup>2</sup> for the trend post implementation is 94.1%. The rate of performance improvement, as measured by cumulative yield, has improved significantly with the implementation of the new PM system.

The production line's ability to control performance is assessed by comparing the variability of the cumulative yield data before and after implementation. The MAD from the model was .00376 prior to implementation and only .000198 afterwards. The standard error of the trend estimate prior to implementation was 4 times the standard error of the error after (.00056 vs .00014). A t-test of the absolute value of the residuals for pre and post TG was significant at the p=.001 level. The conclusion for this portion of the analysis is that control increased significantly with respect to cumulative yield.

# (2) Productivity

A similar analysis was applied to the productivity data for which 27 months of data were available (8 months pre and 19 months post implementation). The same general linear model form was used. Although the overall model is significant at the p=.000 level, none of the individual main or interaction effects are significant. This indicates that there was no statistically difference in either the level or rate of improvement in productivity before and after the implementation of TG. Combining the pre and post data together shows an overall rate of improvement of 19.0 (p=.000) moves per adjusted production FTE was maintained through implementation. The combined data model had an adjusted  $R^2$  of 52.7% and a standard error of 140.6. Figure 3 shows a plot of the data and the regression line.



Fig 3 Graph of productivity data pre and post Top Grading implementation

The variability of the productivity data was also checked for changes using a t-test of the absolute value of the residuals for the data pre and post TG implementation. The residuals were based on the regression for the combined data. The variation was not significantly different at alpha =.1. There was not a significant difference in the level of control of productivity demonstrated before and after TG implementation.

# 5.2 Performance Measure (Independent) Variables

There are definitely changes in the overall, strategic performance variables (cumulative yield and productivity). The question remains as to what part the performance measurement system plays in effecting that change. Although complex relationships invariably exist between variables in a performance measurement system, the relationship between each PM system element and the dependent variables was analyzed individually due to the small sample sizes. A total of 15 performance measures, including yield data for eight different workcentres, were independently analyzed for causality, improvement, and control attributes. The teamwork measure was dropped from the TG system due to adverse reaction from the operators during the first month of implementation. Operators generally did not feel comfortable evaluating their peers, especially with the data being used for individual performance evaluations. The total throughput time for the production line is 2-3 days; therefore, the expectation is that changes to variable will generally be reflected in the current. Summary statistics for the performance measure variables are reported in Table 2. Table 3 includes a summary of the PM system effectiveness analysis results.

Performance Measure						
<u>Variable</u>	<u>N</u>	<u>Mean</u>	<u>Std Dev</u>	<u>SE Mean</u>	<u>Min</u>	<u>Max</u>
Workcentre Yield (2)						
Access	10	0.7989	0.0010	0.0003	0.7969	0.7994
Bump	10	0.7996	0.0002	0.0000	0.7994	0.7998
Connect	10	0.8000	0.0000	0.0000	0.8000	0.8000
Die	10	0.7982	0.0004	0.0001	0.7977	0.7986
Edge	10	0.7999	0.0000	0.0000	0.7998	0.7999
Fab	10	0.7990	0.0008	0.0002	0.7978	0.7997
Grind	10	0.7988	0.0006	0.0001	0.7981	0.7994
Heat	10	0.7979	0.0012	0.0003	0.7964	0.7987
Output	22	3932	493	84	3378	4710
Housekeeping	13	13.15	2.27	0.63	7	15
Employee Suggestions	8	28	18.09	6.39	3	55
Workmanship	13	292	609	169	0	1800
Safety Incidents	13	0.154	0.376	0.104	0	1
Dependability	13	0.538	0.967	0.268	0	3
Total Score	10	56.5	12.38	3.91	38	73

Notes: 1. Values transformed to protect confidentiality.

2. Workcentre names have been disguised.

Table 2 Summary statistics for the performance measures

Performance Measure					I						
Variable Variable	Causality				Improv	/ement	<u>Control</u>				
Correlation Test,	Yield		Productivity		<u>Time</u>		Mean Absolute Deviation		t-test		
Pearson	1	<u>p-value</u>	Ľ	p-value	Ľ	p-value	<u>Early</u>	Late	<u>p-value</u>	Basis	
Workcentre Yield											
Access		ns		ns		ns	1.0E-03	3.7E-04	0.08	mean	
Bump	0.92	0.00	0.66	0.04	0.87	0.00	8.8E-05	4.3E-05	0.09	trend	
Connect		ns		ns		ns	2.1E-05	8.1E-06	ns	mean	
Die	0.93	0.00	0.61	0.06	0.87	0.00	1.3E-04	1.6E-04	ns	trend	
Edge		ns		ns		ns	3.8E-05	1.6E-06	0.04	mean	
Fab	0.98	0.00	0.76	0.01	0.96	0.00	2.3E-04	1.6E-04	ns	trend	
Grind	0.91	0.00	0.88	0.00	0.94	0.00	1.8E-04	4.5E-05	ns	trend	
Heat	0.98	0.00	0.63	0.05	0.88	0.00	5.0E-04	3.7E-04	ns	trend	
Output	-0.43	0.05		ns	-0.56	0.01	130.4	40.6	0.08	plan	
Housekeeping		ns		ns		ns	5.8	0.7	ns	mean	
Employee Suggestions		ns		ns		ns	12.8	16.8	ns	mean	
Workmanship		ns		ns		ns	413.3	443.6	ns	mean	
Safety Incidents	-0.64	0.02		ns	-0.63	0.02	0.334	0.087	0.00	trend	
Dependability	-0.60	0.03		ns	-0.71	0.01	0.748	0.355	0.04	trend	
Total Score		ns		ns		ns	12.3	95	ns	mean	

Notes: 1. P-values > .1 shown as not significant (ns) and the respective statistics are not reported. **Table 3 Summary of PM effectiveness analysis results** 

# (1) Workcentre Yield

There are eight workcentres in this production area that pass material in a serial fashion. Ten months of yield data were analyzed. Five of the eight workcentre yield values (Bump, Die, Fab, Grind, and Heat) are significantly correlated to cumulative yield and have high r values (.91 to .98). This is a direct result of the cumulative yield being an explicit function of the multiplicative product of the serially connected workstation yields. While correlation does not imply causality, the presents of significant correlation means causality cannot be excluded. On the other hand, the absence of significant correlation on the Access, Connect, and Edge workcentres means that their yield PM cannot be causal. The yield at the Connect and Edge workcentres is essentially zero for the period studied. Without variation, causality is precluded. The Access workcentre does have a comparable level mean yield and variation in comparison to the workcentres with significant correlations. The data was also tested to see if a one-month lag between the dependent and independent variables improved the correlations. This would be indicative of a delayed effect in the measurement system. There was no appreciable change in the Access yield PM correlation to cumulative yield.

Workcentre yield was similarly tested for correlation to productivity. The same workcentre yield PM's (Bump, Die, Fab, Grind, and Heat) were significantly correlated to productivity, although at lower r values (.61 to .88) and at higher p-values. Die yield was only significant at the p=.06 level. Being correlated to both strategic performance variables adds support to the use of the yield for these workcentres as PM's. The use of yield for the Access, Connect, and Edge is not supported from a causality perspective. It should also be noted that this analysis does not preclude the possibility that Access yield might have a mediated effect on the dependent variables. A correlation analysis of Access to other workcentre yields did not show any significant effects, but a simple correlation analysis might not defect a relationship due to the small sample size.

Improvement is demonstrated by a positive and significant correlation with time. Workcentre yield was correlated against months of implementation. All of the workcentres showed significant and positive correlations with time, except for Access, Connect, and Edge. In addition, these same workcentre measures failed causality for cumulative yield based on a correlation test. This further brings to question their appropriateness as a performance measures for these workcentres.

Control is demonstrated by a reduction in the variation in the measure for the first five months versus the second five months. A t-test was performed comparing the absolute values of the residuals for each of yield measure that had significant correlations to time and the absolute value of the difference from the means for those that were not significantly correlated to time. All of the MAD's for the workcentre yield PM's decreased over the period studied except for Die. Only Edge was significant at alpha = .05, but that is an artifact of the fact that the only change in yield for that workcentre occurred in the first month of the study. Access and Bump yield demonstrated significant reductions in variability, but only at the alpha = .1 level.

# (2) Output

Output in units was measured at the end of the production line and was a common PM to all teams and operators. Output was actually negatively correlated to cumulative yield and not correlated to productivity with respect to causality. The gross output PM over the period actually decreased while yield increased and productivity as measured by moves per operator increased. This is explained by a shift in product mix over the period from relatively simple products (low moves) to more complex products (high moves). Output remained relatively constant as productivity measured in moves per adjusted production FTE increased. For this measure, performance was increasing, but the

measure was not indicating where credit was due. Raw output is not a good PM with respect to the causality attribute, especially with operators incentivized with extra points to produce "above target".

The problem with using raw output is also demonstrated in the significant and negative correlation with time. Once again, even as the end result productivity was improving, this was not reflected in what should have been a prime indicator of improvement.

Control is demonstrated by a reduction in the mean difference between actual and planned output. For this measure, 22 months of data were available including 4 months prior to TG implementation. A sample of the first 11 months was compared to the subsequent 11 months. Although the mean difference did reduce from 130 to 41 units, it was only significant at p=.08. Since output has a meet or exceed target in the point scoring convention, control in the form of "meeting plan" is not emphasized by the PM system.

# (3) Housekeeping

Material handling and generic housekeeping is measured by a monthly third party audit of the workcentres. Individual workcentre scores were not available, however an average score for the entire production line is used for demonstration purposes. Thirteen months of data were available. The housekeeping PM was not significant with respect to any of the effectiveness attributes. The MAD of the data did decrease from 5.8 to .7, but it was not significant.

# (4) Employee suggestions

An employee suggestion system (ESS) score was a PM system element early in the PM system implementation. Operators were given gold, silver, bronze and recognition awards for suggestions. However, the original ESS was eventually dropped, due to an overly bureaucratic approval system that could not keep up with the load of suggestions in a timely manner. A revised ESS system is currently in place and working well. Only 8 months of data were available for the ESS as an active element in the PM system.

The results for the total employee suggestion PM were very similar to that of housekeeping. None of the effectiveness criteria was significant. The variability of the data actually increased. Suggestions at first increased and then decreased as the system failed.

### (5) Workmanship

Workmanship was measured as the number of rejects for whatever reason that were detected in the customer production area and attributed to Intermediate Assembly. Counts of rejects can be based on whole lot rejections. Thirteen months of workmanship data were available. Evidence of causality and improvement was not present. Variation actually increased slightly over the period, but not significantly. The variation in workmanship is susceptible to lot and run size variations. Whole lots can be rejected if single failures are found. However, the number of months with rejects over the period did increase (2/6 versus 4/7), perhaps reflecting an increased awareness of quality.

#### (6) Safety

Safety was measured as the total number of reportable and non-reportable incidents. Thirteen months of safety data was available. In those 13 months, only 2 safety incidents were reported, so the results for this data should be interpreted with even more caution than allotted to the small sample sizes. Because the 2 incidents occurred in months 1 and 3 of TG implementation, they pick up a significant negative correlation to cumulative yield that improved over the period. However, the correlation to productivity was not significant. Of course, with only 2 safety incidents in the early months, improvement and control show up as significant.

#### (7) Dependability

Dependability was measured based on the number of deviations to established policy on individual absences, late arrival, and other missed work time situations. Thirteen months of data were available. The dependability PM did demonstrate a significant negative correlation to cumulative yield, a significant negative improvement trend, and a significant reduction in variation. However, it also suffers from the same low occurrence problem as safety incidents. Only 6 incidents occurred in the first 3 months and 1 incident in month 9.

#### (8) Total Score

Each of the scorecard elements is translated into point values that are totaled to determine scores for individual operators. The point values and scoring criteria for each of the measures is shown in Figure 1. By selecting the point value assigned to each measure, management can adjust the perceived importance of that measure to overall performance. The total score should be correlated to operation performance, show trends of increasing or decreasing performance with time, and demonstrate reduce variability as individuals understand how to achieve the maximum scores on their performance.

This was not the case for the TG scoring system. An aggregate, average score for the whole production area was developed the test the conditions of PM system effectiveness. The total score was not correlated to either cumulative yield or productivity. This implies that the total score itself was not causing the improved performance. It was also not correlated to time; therefore, it did not demonstrate the ability to show improvement.

Finally, the total score was tested for reduced variability that would be indicative of improved control. The absolute differences between the scores and the mean score for the 10 months for which a complete data set was available was divided into equal early and late groups. A t-test indicated that there was not a significant difference in the means between the two groups. The MAD did decrease from 12.3 to 9.5.

#### 6. Analysis Summary

The implementation of TG resulted in a substantial increase in the rate of improvement and level of control for the strategic variable cumulative yield. Productivity did not change, but continued its positive rate of improvement. Of the PM's used by TG, only the Bump workcentre yield demonstrated all three of the PM system effectiveness criteria (causality, improvement, control) proposed in this paper. The Connect yield, housekeeping, employee suggestions, and workmanship PM's demonstrated none of the effectiveness criteria bringing to question their contribution to strategic performance. Using output as a PM is actually contra indicated due to its negative correlation to yield and time. Finally, the total score, which is intended to be a good overall driver and indicator of performance, is not significant on any of the three effectiveness criteria.

# 7. Conclusion

This paper proposed and illustrated a 2-tier process for analyzing the effectiveness of PM systems. Strategic dependent variables were first identified and analyses for changes related to PM system implementation. Next individual PM's were analyzed to see if they demonstrated the PM system effectiveness attributes of causality, improvement, and control. The analysis process highlighted strengths and weakness of a new PM system implemented in a single production line at HP. In this particular case the overall effect on cumulative yield as a strategic variable was as desired. Seven of the PM's are candidates for elimination from the PM system for being non-causal to either strategic variable. Of the seven PM's remaining only the Die workcentre yield PM does not demonstrate the attribute of improved control as indicated by a reduction in the MAD. The reduction in MAD in all the PM's other than Die yield was problematic from the standpoint of using the PM system scoring to differentiate operator performance for evaluation purposes. The control objective for a PM system tends to reduce variation, while variation reduction as well as improvements to levels of process capability. Also adding to the problem of lack of operator differentiation is the low granularity scoring conventions. For example giving "25 points above, 15 points on, and 0 points below target" eventually causes a string of 25 scorings as process capability is achieved at each new target level. Simplicity in scoring should be traded off against responsiveness.

Future research in the area of PM system design and analysis should address both content and method issues. For example, there is evidence in the literature that PM systems should also be comprehensive [3, 7, 21, 27] and simple [28]. Systematic methods need to be developed to analyze these attributes. There also needs to be an effort to incorporate statistical methods more sensitive to changes present in the small sample sizes of longitudinal data. Future work on this same data set will involve the use of non-parametric data analysis techniques. Finally, the learning that resulted from this study will be used to design the next phase of improvements in IA's PM system. The effectiveness of those improvements will be the subject a future confirmatory analysis.

#### References

- Leong, G.K., D. Snyder, and P. Ward; Research in the process and content of manufacturing strategy, Omega, Int. J. of Mgmt Sci., 18(2): pp109-122, 1990.
- [2] Kaplan, R.S.; Measuring Manufacturing Performance: A New Challenge for Managerial, Accounting Research, The Accounting Review, 58(4): pp686-705, 1983.
- [3] Kaplan, R.S. and D.P. Norton; The Balanced Scorecard Measures That Drive Performance, Harvard Business Review, 70(1): pp71-79, 1992.
- [4] Kaplan, R.S. and D.P. Norton; Putting the balanced scorecard to work, Harvard Business Review, 71(5): pp134-142, 1993.
- [5] Kaplan, R.S. and D.P. Norton; Linking the balanced scorecard to strategy' California Management Review, 39(1): pp53-79, 1996.

- [6] Maskell, B.H.; Performance measurement for world class manufacturing : a model for American companies, Cambridge, Mass, Productivity Press, p20, 1991.
- [7] Kaydos, W.J., Operational performance measurement: increasing total productivity; Boca Raton, FL: St. Lucie Press, 1999.
- [8] Chang, R.Y. and M.W. Morgan; Performance scorecards: measuring the right things in the real world, 1st ed., San Francisco, Jossey-Bass, 2000.
- [9] Tesoro, F. and J. Tootson; Implementing global performance measurement systems, a cookbook approach, Jossey-Bass/Pfeiffer: San Francisco, p10, 2000.
- [10] Norton, D.P.; Measuring Performance in the Organization of the Future, Nolan Norton Institute, Lexington, MA, 1991.
- [11] Blenkinsop, S.A. and N. Burns; Performance Measurement Revisited, International Journal of Operations & Production Management, 12(10), pp16-25, 1992.
- [12] Meyer, C.; Fast cycle time : how to align purpose, strategy, and structure for speed, New York, Toronto, Free Press, Maxwell Macmillan, Canada, Maxwell Macmillan International, pp234, 1993.
- [13] Stalk, G. and T.M. Hout; Competing against time : how time based competition is reshaping global markets, New York, London, Free Press, Collier Macmillan, 1990.
- [14] Chang, R.Y. and P. De Young; Measuring organizational improvement impact : a practical guide to successfully linking organizational improvement measures, Quality improvement series, Irvine, CA, Richard Chang Associates, pp9, 11, 1995.
- [15] Hayes, R.H., S.C. Wheelwright, and K.B. Clark; Dynamic manufacturing: creating the learning organization, New York, London, Free Press, Collier Macmillan, 1988.
- [16] Lynch, R.L. and K.F. Cross; Measure up! : yardsticks for continuous improvement, 2nd ed., Cambridge, Mass.: Blackwell Business, 1995.
- [17] Hall, R.W.; Zero inventories, ed. A.P.a.I.C. Society, Homewood, IL, Business One Irwin, 1983.
- [18] Dixon, J.R., A.J. Nanni, and T.E. Vollmann; The new performance challenge: measuring operations for worldclass competition, Homewood, IL, Dow Jones-Irwin, 1990.
- [19] Dow, D., D. Samson, and S. Ford; Exploding the myth: do all quality management practices contribute to superior quality performance?, Production and Operations Management, 8(1), pp1-27, 1999.
- [20] Grady, M.W.; Performance Measurement: Implementing Strategy, Management Accounting, 72(12): pp49-53, 1991.
- [21] McDougall, D.C.; Learning the ropes: how to tell when you've found an effective performance measurement system for manufacturing. Operations Management Review, Fall 1987 & Winter 1988, pp38-48, 1988.
- [22] Deming, W.E.; Quality, productivity, and competitive position, Cambridge, MA, Massachusetts Institute of Technology Center for Advanced Engineering Study, vii, 373, 1982.
- [23] Crosby, P.B.; Quality is Free, New York, N.Y., McGraw-Hill Book Company, 1979.
- [24] Globerson, A., S. Globerson, and J. Frampton; You can't manage what you don't measure: control and evaluation in organizations, Aldershot, Hants, England; Brookfield, Vt., USA, Gower Pub, 1991.
- [25] Cohen, J. and P. Cohen; Applied multiple regression/correlation analysis for the behavioral sciences, 2nd ed., Hillsdale, NJ, L. Erlbaum Associates, 1983.
- [26] Lim, P. and S.L. Teo; Re-designing for Fast Manufacturing, in Manufacturing Management Symposium XIII, Hewlett Packard, 2001.
- [27] Lean Manufacturing Measurables, Ford Motor Company Lean Resource Center, 1999.
- [28] Olsen, E.O. and P.T. Ward. Performance measurement system simplicity, under review for Decision Sciences Institute Conference, San Diego, CA, 2002.