LESSONS LEARNT IN DEVELOPING EFFECTIVE PERFORMANCE MEASURES FOR CONSTRUCTION SAFETY MANAGEMENT

Marton Marosszeky¹, Khalid Karim², Steven Davis³, Nitin Naik⁴

ABSTRACT

This paper presents the lessons learned to date in a safety-performance benchmarking project, where the client funded research to develop measures that would drive improvement on two concurrent hospital construction projects.

The study shows the development of the performance measurement regime that was adopted and the complexity involved in developing effective feedback mechanisms for supervisors and workers on site.

This work is still in progress and each week the research team and the project team gain new insights into the difficulties that are faced in any attempt to transform the construction workplace.

The process to date has been crudely modelled, however it has to be recognised that such models are not generic, rather they reflect the particular process on a project.

KEY WORDS

Safety in construction, performance measurement, feedback

¹ Multiplex Professor of Engineering Construction Innovation, Director ACCI, School of Civil and Environmental Engineering, University of New South Wales, Australia.
² Deputy Director ACCI, School of Civil and Environmental Engineering, University of New South Wales, Australia.
³ Lecturer, School of Civil and Environmental Engineering, University of New South Wales, Australia.
⁴ Postgraduate research student, School of Civil and Environmental Engineering, University of New South Wales, Australia.
INTRODUCTION

Safety and quality continue to remain critical priorities in the context of improving productivity and efficiency in the construction industry within Australia as well as overseas. Larsson and Field (2002), in their analysis of the Victorian construction industry, provide evidence of continued unacceptable risk exposure in terms of safety. Edwards and Nicholas (2002) in their study of the UK construction industry portray it as the most hazardous industry. Similarly, recent studies by the authors have showed that the cost of quality related problems is of the same magnitude as the profitability of organisations in the sector. (Marosszeky et al. 2002, Thomas et al. 2002).

In the project reported in this paper the client, a regional hospital authority, identified safety and quality performance as critical areas for improvement. The client had two projects of a similar scale (> $100 million) being constructed concurrently and decided to fund the research with a view to drive process improvement through comparison between the two projects. Both projects were to be built by the same general contractor and the client had their agreement to cooperate in the project and implement ideas that emerged out of it. While the original intent had been to conduct performance measurement in both safety and quality, the lack of formal management processes in relation to quality made this task impractical and this paper presents the lessons learned in relation to safety performance measurement and benchmarking.

The research method used was essentially an iterative process that involves study/analysis of the subject of measurement (e.g. process), identifying potential performance measures, prioritising and accepting/discarding measures, and finally the development/refinement of Key Performance Indicators, as well as the feedback mechanism, and their implementation. Letza (1996) formulated a flow chart for this iterative mechanism as discussed later.

While the work is still in progress, a number of lessons have been learned in the process – some of them surprising. This paper presents those lessons by placing the current project in the context of past research, followed by a brief description of the research objectives and methodology, and then providing a detailed description of how the process unfolded.

PERFORMANCE MEASUREMENT AND FEEDBACK IN SAFETY

Safety management systems have largely been developed in response to statutory requirements. Thus reporting has focused mostly on mandatory information related to accidents and injuries. Such measures suffer from three drawbacks. Firstly, they measure what happens after the event and are reactive in terms of management response. Secondly, in the absence of any proactive measure, causal relationships cannot be established. Thirdly, they are negative in nature and acknowledged as being unsuccessful as measures of safety performance (Trethewy et al. 2000, Mohamed 2003).

More recently, the focus is shifting towards more detailed management oriented measurements that have the potential to influence processes on the project being assessed. These include the subjective performance rating used by Jaselkis (1996) and development of the Site Safety Meter (based on a traditional site inspection) by Trethewy et al. (2000). Marsh et al. (1995) used measures such as access to heights, housekeeping, and personal protective
equipment to try to influence behaviour and Mohamed (2003) formulated a performance measurement system at an organisational level that did suggest the use of operational (i.e. site based) performance measures, this approach however was not supported in responses from industry.

The discussion and development of performance measurement in the 'Lean Movement' in construction can be classified into two broad categories. One strand, somewhat consistent with preceding attempts, mostly describes performance in terms of outcomes, the other focuses on processes.

Typical examples of outcome based performance measurement relate to safety, quality and environmental failures, productivity, reliability of deliveries, customer satisfaction, cost and schedule variations, design/documentation deficiency, and management dimensions such as leadership and training (Ellis Jr., 1997; Gaarslev, 1997; Tilley et al., 1997; Ghio, 1997; El-Mashaleh et al., 2001; Saurin et al., 2001; Alarcon et al., 2001; Ramirez et al., 2003). On the other hand, performance measurement in relation to processes includes waste as defined in Lean Construction (Alarcon, 1997), look ahead planning and plan percent complete (PPC) (Ballard, 1997; Ballard & Howell, 1997), safety process improvement (Saurin et al., 2002), quality process improvement (Marosszeky et al., 2002; Thomas et al., 2002) and measuring a firm's conformance to lean ideas (Diekmann et al., 2003).

In much of the literature on construction performance measurement the underlying assumption has been that simple, straightforward feedback will serve the objective of continuous improvement. Thus feedback has been informal and un-structured despite recognition of the fact that a learning culture is critical to continuous improvement (Scott and Harris, 1998; Loo, 2003). Consequently in construction, scant attention has been paid to the actual feedback mechanism to ensure that it is effective, or to the questions that need to be asked such as:
- Exactly what do the numbers mean to the recipient of information?
- What information is most useful?
- What should be the format of presentation? and
- At what level should the information be initially fed for further dissemination?

These need to be considered in order to avoid the possibility of different operators associating different meaning to the same information, or worse still, the feedback not having the impact that it warrants.

It is quite possible, as became evident in the work being reported in this paper, that what is deemed to be positive feedback by one party is seen as negative by another. Santos and Powell (2001) found that this even occurred where an information recipient was a stakeholder in the performance measurement development process itself. There may be several reasons for this, one is that honest criticism is hard to take (Cleeton, 1992), another factor may be that the information recipient feels threatened by the information or its impact. Unfortunately such reactions generate resistance to change (Santos and Powell, 2001) and can undermine a process of continuous improvement, the objective of the performance measurement in the first place.

Santos and Powell (2001) discuss this issue in terms of push and pull learning. Push learning involves putting external agents such as researchers or consultants in charge of deciding what the learners need to learn. In contrast, in pull learning individuals in the
enterprise are in charge of the learning process and its objectives. Pull learning is much more likely to create a ‘learning mood’ that maximises the acceptance of feedback. However, push learning can be the trigger for stimulating pull learning, and this should be the objective of push learning.

Forza and Salvador (2000) investigated performance measurement in terms of the “distinctive dimensions” over which the quality of the feedback can vary. They also sought to determine whether these distinctive dimensions change as the hierarchical level of the information user changes. Two of these dimensions were found to be useful for the detailed operational performance feedback that is the focus of this study. These are *relevance as performance feedback orientation to the achievement of objectives* and *dynamic adjustment of performance feedback*.

The former relates to elements that show objectives and trends, the relation of the measures to planned objectives and the timely receipt of information for the initiation of improvement actions. The latter deals with the changes to indicators depending on the situation being addressed, and to reflect change in planned objectives and programs.

**RESEARCH OBJECTIVE AND METHODOLOGY**

The objective of this study was to investigate the development and implementation of safety related performance measurement at the detailed operational level within the supply chain, i.e. the work by trade subcontractors, and identify an effective feedback mechanism to stimulate learning and improvement. It focuses on the overall process environment, as opposed to the behaviour of individuals (Behaviour Based Management), the latter deserving attention in its own right.

The general framework used to measure performance is presented in Karim et al. (2003). In brief it involves three parts. Firstly, is management complying with the management system? This usually involves following plans and looking for problems. Secondly, is management responding to those problems? Thirdly, is this improving the project? This requires outcome measures. However, the outcomes should be measured at the process level rather than the project level to enable feedback.

The skill of effective performance measurement is in several areas, first of all it is important to identify areas that reflect critical business or process goals so that the measurement task reflects strategic thinking, secondly specific measures have to be identified that are easy to measure, and thirdly feedback must be designed so that it creates the desire in those measured to improve on past performance.

The development of a performance measurement framework is essentially an iterative process that involves study, experience and analysis as well as negotiation with operational management at every stage of the process. Potential performance measures were evaluated based on past experience and project goals; they were then prioritised and selected based on the availability of information on the project and compatibility with existing management processes and philosophies. The difficulty within these negotiations should not be underestimated, even when senior management of the constructor is fully behind the project getting buy in from site staff can be difficult and slow.

Once the overall framework is agreed, implementation is still ahead and there is a great deal yet to go; this includes developing measurement protocols, trialing and refining those
with operational management, developing feedback for discussion and review, and then refining and implementing the feedback until it is in a form that operational management accepts.

Letza (1996) formulated a flow chart for this iterative mechanism as shown in Figure 1a below. In Figure 1b the actual process undertaken on this project is summarised. The essential difference between the two models is that there are several stages at which agreements have to be negotiated, and approvals secured. As the detail of the proposed performance process develops new issues emerge, often requiring renegotiation. In addition, at each of these stages there is a need for significant development and testing of ideas, refinement, acceptance and agreement.

![Diagram](image)

**Figure 1: Iterative process for developing key performance measures**

**PROCESS DESCRIPTION**

The process of developing the framework essentially began with documentation analysis to examine previous experience with performance measurement within the area of safety. A preliminary framework for performance measurement was then designed, reviewed and refined through an iterative process of joint stakeholder meetings from both projects and the concurrent development of ideas. Data collection commenced once this was agreed. The overall initial framework is summarised in Table 1.

There was a quick realisation that in some cases these scores were simplistic in that it was possible for both sites to achieve scores consistently close to 100%, even though normal
industry practice in these areas is much lower. So for example, all targeted toolbox meetings were held. At first both sites found it difficult to achieve the number of planned audits, however once the monthly performance scores were published the actual quickly converged with the target. This supported the adage that if you measure performance, it will tend to improve. This initial framework was then subsequently developed and modified over the following months along the basis of the process described in Figure 1. From the outset, as exemplified in measure 5, we were looking to compare the performance of subcontractors in order to generate some healthy competition and to create the basis of public recognition of the best subcontractor each month.

The following two examples describe the application of the framework in the elimination of safety hazards on each site. The first is an examination of how effectively subcontractor management was responding to hazards identified in site safety walks, and the second is in the use of performance measures arising out of the application of the Site Safety Meter (SSM) (Trethewy et al, 2000). It is pertinent to note here that, all along, some contractor staff viewed the SSM with a degree of suspicion and so initially it was only used on the leading deck, with more widespread use occurring slowly.

The safety walk is one of the major safety management processes used on construction sites in Australia to control safety. In this walk the site safety committee systematically walks through the whole site and documents every unsafe working practice and hazard they see. Furthermore, they identify potential upcoming hazards to alert the workforce to them in advance of the work being done - an issue that caused us some difficulty mid-way through implementation. Each issue listed is assigned to a specific party to rectify, usually the subcontractor responsible for the worker or area or, in some cases, the head contractor. The party assigned is then expected to fix the problem or take precautionary measures as quickly as possible and to report back to the safety committee a week later at the following meeting.

Initially we applied the second part of the framework (management responsiveness) to this by determining the number of errors identified in the safety walk closed out within one week. This turned out to be unsatisfactory because 100% of items were closed out, yet we observed very similar items appearing on the list the next week indicating a lack of learning. This led us to measure the number of hazards that are being created by each subcontractor. Table 1 illustrates a typical result from one of the sites. The number of errors per subcontractor was divided by the work hours to make the results comparable. The bar chart

<table>
<thead>
<tr>
<th>No.</th>
<th>Performance Measure</th>
<th>Outcome Indicator</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Toolbox Talk</td>
<td>Planned vs. Actual - %</td>
<td>The first measure was simply to record whether the planned process of at least 1 toolbox talk per week was achieved. Toolbox talks include the weekly safety committee meetings. The second measure was to record the % of safety issues that had been rectified within the week.</td>
</tr>
<tr>
<td>2</td>
<td>Safety Walk</td>
<td>Closed out items - %</td>
<td>Safety walks generate list of hazards that have to be rectified, the % of items closed out by the end of the week 3 days later was to be recorded.</td>
</tr>
<tr>
<td>3</td>
<td>Task Observation</td>
<td>Planned vs. Actual - No.</td>
<td>The contractor planned to conduct a certain number of task observations each week to assess whether the subcontractors were working according to their own documented plans. The measure was to calculate the actual number of planned audits that was executed and whether these identified as hazardous were actually recorded.</td>
</tr>
<tr>
<td>4</td>
<td>Site Safety Meter</td>
<td>Score based on site inspection on site parameters related to OHS</td>
<td>The site safety meter (SSM) was used to score the safety environment, because of scepticism about the measure initially, it was only used on the leading deck. This generated a score for each site.</td>
</tr>
<tr>
<td>5</td>
<td>Subcontractor Safety System</td>
<td>This is a composite score as a %</td>
<td>Initially the score was based on a quarterly audit of subcontractor safety systems monitored by the report of on-site audits of construction activities undertaken at the main site. After a period of negotiation it was agreed that the measure was to be redefined as follows: awarded to subcontractor once the monthly period rather than a cut off audit.</td>
</tr>
<tr>
<td>6</td>
<td>Knowledge &amp;</td>
<td>Knowledge improved - %</td>
<td>A questionnaire was developed to assess worker knowledge in relation to safety and it was repeated at the next walk 6 months to see whether overall knowledge about safety issues improved.</td>
</tr>
<tr>
<td>7</td>
<td>LTPs, MTIs &amp;</td>
<td>Average lost time</td>
<td>Mandate safety performance statistics were to be collected and collated for comparison with the other measures.</td>
</tr>
</tbody>
</table>

Table 1: The initial performance measurement framework
then clearly indicated where the priorities needed to be assigned for targeting specific trades for improvement.

However, the actual question that we wanted to answer was whether the trade groups were learning from their mistakes, or were they just remedying specific errors and repeating them week after week? Our position was that repetition indicated the absence of learning and we wanted to see how much learning was happening. This led us to take an interest in the percentage of items found on each safety walk that had occurred previously (Figure 3).

It quickly became obvious that the issue of repetition is complicated and two specific problems emerged. The first is that if you define repetition as the same unsafe practice or hazard reoccurring then this percentage will naturally increase as the project progresses as there is a much higher probability that any given issue has arisen before, leading to almost every error becoming a repeat item and the measure losing its sensitivity. This led us to define repetition as an item that had occurred in a safety walk in the immediately previous month. This also had the advantage that the time over which the repetition was based was limited and so scores from earlier in the project could be compared with those later in the project.

The second problem was that as we probed more deeply and spoke of publishing data on the walls of the site sheds, everyone started to look more closely at the scores and it was realised that some of the issues listed on safety walks are potential hazards in work tasks yet to be done and hence they should not be included. This led us to more carefully analyse the listed items.

The next problem was how to present the data. Our initial presentation showed the number of items for each trade so that a comparison between trades could be made. This presentation also showed changes over time, month by month. However, the head contractor on the site rightly pointed out that these measures are of limited value unless the data shows the breakdown of the individual problems so that the subcontractors know where to focus their training.
Consequently, a presentation showing the number of individual items and changes over time was created (Figure 4). The head contractor found this presentation to be extremely useful. The next issue was the need for consistency between the safety committee minutes and the graph, for example if the minutes mentioned that a "general cleanup was required" then it would be inappropriate if the graphical presentation referred to this as "dirty work area". The requirement for accuracy is critical if information is being used to recognize excellence and to promote friendly competition between the subcontractors on the site.

A second application of the framework was to use the SSM to focus efforts on improvement of the physical safety environment. This process scores a range of potential hazards on a site that correlate with the highest safety risks (as a percent correct) and gives a simple score, enabling the safety of the project environment to be compared week by week. On this project the categories shown in Table 2 were used.

The site is divided into areas and, in each area, each item is rated as acceptable or unacceptable. The score is calculated as the number of acceptable items divided by the total number of items and given as a percentage. At first an overall score was generated for each site and the trends were plotted. It was found that relative to previous uses of the SSM on other sites the safety environment on both projects rated well. In order to identify the weakest aspects of site safety, the focus shifted to a closer examination of errors, this was examined both by trade and by issue.

Some site managers were reluctant to use the SSM. Two reasons were given, one was that if the site was to be surveyed at random times, there was a possibility that the data might be gathered on a day when the site just happened to be in an unusual state of disarray. A second issue of concern was that positives were counted in calculating the SSM measure rather than simply listing the errors. This is a conceptual problem. It seems that some managers are comfortable with the traditional approach to safety management and the outputs that it yields, even though they do not engender a move towards continuous improvement. In general, management seemed to be against any form of feedback that could be regarded as criticism.

However, it was finally agreed that the collection of data could occur during the safety walk.

---

**Table 2: Site Safety Meter (SSM) Category**

1. Working Habits - Use of protective gear and risk factor.
2. Order & Tidiness (Housekeeping) - Waste bins, work area tidiness, access ways.
3. Electrical & Lighting - Temporary electrical boards, lights and tools. Lighting to work area.
4. Scaffold and Ladders - Erected and secured correctly.
5. Protection Against Falls and Falling Objects - Perimeter handrail, penetrations, overhead protection.
6. Plant and Equipment - Hoist or crane, concrete pump, jackhammer, and other.
The next idea regarding the implementation of the site safety meter related to the rate of improvement. The purpose behind this was the need to use it as an incentive for promoting healthy competition. It emerged that rewarding on the basis of an improvement from 80% to 90% was similar to an improvement from 98% to 99% in that in both cases the number of errors was halved (in actual fact the latter is more difficult to achieve but this was ignored to avoid complication). However, in the original mechanism for use of site safety meter this was not so obvious and a subcontractor starting with a poor value appeared to make the most improvement. In order to overcome this in presenting the data, we decided to present it as the number of errors divided by the total (Figure 5) rather than as the number of correct items divided by the total. This changes the above to examples to 20% to 10% and 2% to 1% and makes it more obvious that both have halved their error rate. However, 10% wrong sounds worse than 90% right and so this change was perceived as negative as well, further underscoring the sensitivity of site management to anything that could be construed as criticism.

Another issue discovered during the application of site safety meter was that as work changes, and the crew moves from safer work to more hazardous work, the same crew may create more hazards in their work. However, this may not be reflected in the safety meter score since the computation depends on total number of activities.

CONCLUSION
The research had started with the objective of detailed process level performance measurement and learning about effective feedback mechanisms, but even then the researchers were surprised by the way the process unfolded. It is evident now that it's a different story when trying to stimulate improvement at a detailed process level. Difficulties will be encountered. Even the stakeholders (senior site management) involved in the process of developing and refining performance measures don’t see the problems until feedback starts. As a result, generalised models such as that proposed by Letza (1996) understate the complexity of the task. For this reason, design and piloting of feedback should be concurrent with development of performance indicators as far as possible. Similarly, while the feedback dimensions noted previously from the work by Forza and Salvador (2000) were expected to have a bearing, they took on a different meaning on this project. In their case dynamic adjustment of performance feedback relates to changing organisational objectives, whereas in this study the objectives within the process also changed – e.g. changing the objective from positive to negative feedback or vice versa. Similarly “relevance” as defined by them relates to optimum decision making whereas in this case it related to delivering the appropriate
“message”. Consequently, a number of observations emerged during the study. These include:

- Some production process anomalies become apparent only when detailed feedback is developed.

- Once communication of detailed feedback starts, there is a desire to filter information and make it more targeted. This in turn leads to changes in the type and format of presentation. Even the choice of words becomes an issue in detailed feedback.

- Detailed process level performance measurement and feedback encounters the same type of resistance and suspicion that existed at the management level when non-traditional performance measurement was first introduced. In this case, even the timing of inspections was an issue for a while. Strong leadership and skilled negotiation are prerequisites for success.

- In the drive to improve process reliability, there is no substitute for getting into the specific detail of the errors that are to be avoided. While at a project level, say safety environment of the entire site, broad measures raise awareness and lead to overall improved performance, it was found that even high aggregate scores can hide specific areas of weakness.

The purpose of measuring performance is to create feedback that will lead to improvement. It is relatively easy to gather lots of performance data, however it takes a great deal of effort to extract from the data useful trends and to identify where efforts for improvement should be directed. Finally, it is surprisingly difficult to develop ways to present the data so that the information leaps out in a way that will create a reaction that leads to improved performance.

The process described in this paper is currently being tested by implementation on site. On completion of the trial, techniques such as time series analysis, linear regression, and multiple regression will be used to examine the relationship between the framework and improvements in safety. Furthermore, a survey instrument will be used to determine whether some of the performance measures are more effective than others. Pending this detailed analysis, anecdotal evidence suggests that the management of safety on the projects under study is improving as a result of this work.

Acknowledgement: The authors gratefully acknowledge the funding and in-kind support for this research provided by the Central Coast Area Health Service of New South Wales, Australia. The authors are also grateful to the Walter Construction Group for their significant contribution of time and resources to the project.

REFERENCES


