

VERIFYING THE VALUE OF OBJECTIVE MEASURES

A Proposal for a Systematic Evaluation of Measures

Harald Kjellin

Department of Mathematics and Science, Kristianstad University, 291 88 Kristianstad

Keywords: Performance Metrics, Designing Measures.

Abstract: The results of work in any section of an enterprise should preferably be described in a way that makes the results suited for benchmarking with other sections of the enterprise. The same goes for individual work results. Results are easily compared if they are measured according to some numerical standard. Numerical measures can be generalized and standardized until they can be considered as having a high degree of "reusability". There are several types of enterprise models that include the use of reusable "soft" numerical values. With "soft" numerical values I refer to the type of values that cannot be directly measured in relation to objective facts but are artificially constructed measures that includes some kind of subjective estimation for calculating the value. Another requirement on such measures is that it should be possible to use them for comparing performance between individuals or between units of an organization or between organizations. These measures can, for instance, be used for customer appreciation of their relationships with the organization, as is often recommended in the method called "Balanced Scorecards" or they can be used when giving students numerical values as credits (points) for passing university courses. A summary of informal evaluations is presented. The evaluations concern how "soft" measures have been implemented in organizations. The results of the evaluations show that objective values based on facts can be combined with subjective estimations in a way that makes them less vulnerable to people manipulating the measures and less vulnerable to the subjectivity of superiors when estimating the quality of the results.

1 INTRODUCTION

When people discuss the advantages of using ontology they may claim that: a) It facilitates communication within the organization, or b) Ontology makes it possible to automate parts of the communication, or c) Ontology makes it easier to communicate about complex phenomenon, or d) Ontology facilitates the detection of possible misunderstandings (Ushold & Gruninger, 1996) Such claims are similar with the claims stated for other types of methods for structuring information or representing knowledge. For instance in the area of Artificial Intelligence it is well known that knowledge must be represented in a way that makes it possible to structure knowledge in a coherent and logical way or it will not be possible to process the knowledge automatically. Another similar type of claim is that if Conceptual Analysis and Conceptual Modelling are carried out the right way it is possible to create any kind of relational database, with the conceptual model as a specification, since the correctness of the conceptual model will guarantee

that the database will work according to theory. A third example of the need for formalizing relationships in information structures can be found in the area of Knowledge Acquisition where, for instance, it is important that knowledge is structured according to the various perspectives you have on the knowledge when you are solving a specific task.

There is presently many areas of science that focus on the metrics of the output value of phenomena. In the area of "usability metrics" as described in (Nielsen, 1992) we find many intricate ways in how to use heuristic measures, but they differ from the proposed approach in that they do not consider the measures as a standard for continuous use by end users. In the area of balanced scorecards we can see the continuous daily use of performance measures (Kaplan, 2005) similar to the ones proposed here. A slight difference is that we propose measures with a focus on defining the production in itself. Making a measure the reference that people talk about when discussing the output from work. Another similarity can be found with the type of

employee reward proposed in (Armstrong, 1999). In our examples we show how the reward approach can be carried out further towards a definition of the output of the work where the measure and the output is the same thing. There are different theoretical approaches to measurements described in (Stevens, 1946), (Lorge, 1967), (McGonagle & Vella, 1990), (Miller, 1991). There is, however, no specific theory in the above that is utilized as a base for this paper. Instead we have a general approach when discussing empirical evidence concerning how various theories and principles about measurements can be combined in order to facilitate communication about what is valuable and useful in the work place

1.1 Measuring Performance

When large and/or complex information systems are designed it is often necessary to reduce the complexity of the information by establishing unambiguous definitions which may secure that people perceive the information from the same perspective when it is communicated (Chandler, 1992). Such definitions have the same function as the kind of "meta-communication" that is needed to establish an agreement about how the information should be communicated in a dialogue between two persons. The definitions of relationships in communication structures can also be seen as a protocol or grammar for how to interpret messages. A problem with defining the logical structure of: ontology, taxonomy, legend, knowledge base or grammar is that the definition of the structure in itself may be difficult to communicate and it may also be difficult to implement. Apart from this it may create organizational problems when it is enforced within an organization (Argyris, 1991). People may not understand the necessity of the imposed standard and may not be willing to conform to ideas created by someone who they have no relationship with.

Human beings need long-term agreement for how to standardize communication in order for the communication to function efficiently. The problem with this is that the creation of communication standards requires extra resources. The cost of implementing and managing the standards may be larger than the benefits from using them. Another problem with using standards is that they inevitably cause a reduction in the richness of the communication (Nonaka, 1994). There are examples of the exception handling being more expensive to manage than the management of the whole system, which shows that a strict enforced structure may create an efficient communication as long as all

needed variations of the communication can be preconceived, but when it cannot, the exception handling takes time.

Some standards for quantifying the value of results are known as being successful. An example is the often recommended standards used in "Balanced Scorecards" where measures are created for appreciating the value of customers' relationships with the organization. Another example of a successfully implemented standard is the convention to give students numerical values as credits (points) for passing university courses. In Sweden it is very important that such credits follow the national standard in order to allow students to combine studies from different universities. What seems to be common for successful standards is that they combine the recording of facts with some kind of subjective estimation of quality into a numerical value.

1.2 Creating Measures

As this paper only reports a summary of findings I will only exemplify the steps in the development of measures. Initially questions were asked concerning the employees attitude to the results of their work:

- How do you want to be recognized in your work?
- When do you enjoy your work?
- When do you feel that you have produced something valuable?
- Who will benefit from your work.

1.3 Eliciting Measures Via Interviews

Once the value of the results had been specified the interviews continued step by step to gradually establish a measure that was considered as being as objective and measurable as possible. The following are examples of questions that were used in the interviews:

- What could you say to your colleague or superior to prove that you really have achieved something?
- What kind of factual results could be used to compare the results between employees?
- Is it easy to estimate the quality of the results after the quantity of the results has been determined?
- Is there any kind of results related to the value X that can be measured?

1.4 Testing and Modifying the Measures

After implementing the measures they were tested. In some cases an extensive test was carried out, but in most implementations this was not feasible since it would take a very long time to secure reliable results. The major way of finding evidence of the usefulness of the measures was to ask questions to managers and employees of organizations. The questions concerned the usefulness of the measures and they can be exemplified by the following:

For the approach in general:

After having tested these measures. Do you believe that the approach can be used to improve the competitiveness of the organization?

- Do you think that other organizations will be interested in creating similar measures in the near future?
- Would you advice the general management to use these or similar measures in the organization?
- Can the measures be used for benchmarking the performance within the organization?

Specific aspects of the approach:

- Do you believe that the company will have any use of this measure?
- Should the measure be implemented on a large scale in the organization?
- How should it be modified to be more useful?
- Do you see any difficulties with motivating employees to always document their results according to the instructions?

2 WHERE MEASURES WERE USED

This article should be seen as a position statement that gives an overview of the experienced usefulness of measures in various environments:

Performance measures for students in the form of taken credits. 14 years of experiments showed that it is possible to use formal quantifications as a base for evaluating the quality of student performance in their thesis writings and shorter assignment writings. The investigations showed that

the students appreciated to get a distinct numerical grading.

Descriptions of priorities of activities in the area of Knowledge Management. 7 years of tests in smaller companies in the vicinity of Stockholm showed that employees in general strongly disagreed with trying to represent crucial information and personally acquired knowledge according to a specified ontology. At the same time most of the employees were positive to the creation of measures they could use to prove that they had reached their targets in their work. These results were also described in (Kjellin, 2002).

Agreements about principles in negotiations within media organizations. During 5 years there has been cooperation, with the national society of journalist in Sweden, about how to measure the results of their work and how to claim reward for creative work done outside of the work specifications from their employer. Similar results had also been presented in (Armstrong, 1990). The major benefit from the investigation was claims from the journalists that the measures makes it much easier to reach agreements with their employers who dared to delegate responsibility to the journalists since they could trust that the company would only pay for deliverances of high quality results.

The decision support information concerning the values of outsourcing. Managers involved in outsourcing were testing measures for securing the productivity of their personnel. The major conclusion was that the largest benefit from outsourcing was the enforced formalization of measures for costs and benefits from the outsourcing which made later reorganizations easier.

Board evaluations with a focus on the incentive systems. During 3 years 120 measures for evaluating the performance of corporate management was developed and tested on large companies in Sweden. The history behind the measures was similar to how the Sarbanes-Oxley act was developed in the US as a result of the crisis in the companies like Enron, World-Com, etc. The most appreciated feature of these measures was that they secured the objectivity of the measured performance and also secured that the measures could not be manipulated by powerful board members who were sensitive to any type of analysis of their behaviour.

2.1 A Summary of Examples

When all examples were analyzed it could be seen that managers in Swedish organizations:

- are determined to increase their use of soft numerical values
- believe that there is an increased need for soft numerical values in corporations
- see the use of outsourcing as a means to create reliable soft numerical values
- see the use of soft numerical measures as a way to delegate power and decentralize organizations without losing the control that is needed to coordinate the organization.

I also found that:

We can recommend people who create artificial measures to avoid measures that are solely based on facts that can be manipulated. As the measures are used in a competitive environment it is important that the measures cannot be used to manipulate results to create individual advantages. When a part of the measuring is related to a subjective estimation of the quality of the results it is easier to establish a useful numerical measure. The quantitative part of the measure secures that the evaluation is sound and efficient while the qualitative subjective part secures that results cannot be manipulated. This second part can be efficient since it does not include a control of details.

A great reward from evaluating results in relation to costs on a detailed level is that this enforces a creation of standard numerical measures. This can, for instance, be seen in companies who have been engaged in outsourcing. The managers in these companies often realize that a considerable part of the benefits from the outsourcing is that it facilitated a "bottom-up" reorganization of the company based on the formalization of measures of results.

2.2 Epilogue

Objective values based on facts to measure quantitative results can be combined with estimations of the quality of the results in a way that makes them: 1) less vulnerable to people manipulating the measures, and 2) less vulnerable to the subjectivity of superiors when estimating results. In cases when the measures were only based on subjective evaluations the persons whose performance were measured often felt that the measures were unfair and that they were erroneously evaluated by their superiors

2.3 Future research

I am presently looking for companies who are willing to let me implement and test the use of measures on a larger scale. The final goal is to create ontology of measures for some branch of industry. I believe that such ontology could be used for creating a more competitive industry and an industry that would be very skilled in knowledge transfer, knowledge refinement and outsourcing.

REFERENCES

- Argyris C., 1991. Teaching Smart People How to Learn. *Harvard Business Review* 69(3) 99-109.
- Armstrong M., 1999. *Employee Reward* London: CIPD
- Chandler, A. D., 1992. "Organizational Capabilities and the Economic History of the Industrial Enterprise." *Journal of Economic Perspectives*, vol. 6, no. 3, pp. 79-100.
- Kaplan R., 2005. "Add a Customer Profitability Metric to Your Balanced Scorecard," *Balanced Scorecard Report*, Vol. 7, No. 4, July-August 2005.
- Kjellin, H., 2002. How Active Learning Motivates Employees to Share Knowledge, *The European Conference on Knowledge Management*, Trinity College, Dublin, Ireland.
- Lorge, I., 1967. *The Fundamental Nature of Measurement*. In D.N.Jackson, & S. Messick (Eds.), *Problems in Human Assessment*. New York: McGraw Hill.
- McGonagle J.J. & Vella C.M., 1990. *Outsmarting the Competition: Practical Approaches to Finding and Using Competitive Information*. Naperville, IL: Sourcebooks.
- Miller, D.C., 1991. *Selected Sociometric Scales and Indexes*. In *Handbook of Research Design and Social Measurement*. 5th ed. Newbury Park, CA: Sage.
- Nielsen, J., 1992. Finding usability problems through heuristic evaluation, *Proc. ACM CHI'92*, Monterey, CA, 3-7 May. 373-380.
- Nonaka, I., 1994. "A Dynamic Theory of Organizational Knowledge Creation". *Organization Science*. Vol.5, No.1. pp. 14-37.
- Stevens, S.S., 1946. *On the Theory of Scales of Measurement*. *Science*, 103, 677-680.
- Uschold, M., Grüninger, M., 1996 *Ontologies: Principles, Methods and Applications*, *Knowledge Engineering Review*.