VALIDITY PROBLEM IN PERFORMANCE MEASUREMENT: A CONCEPTUAL ANALYSIS WITH AN ILLUSTRATIVE CASE ON QUALITY COST MEASUREMENT

Petri Suomala, Mikko Varila and Ilkka Jokioinen
Cost Management Center
Tampere University of Technology
Institute of Industrial Management
P.O. Box 541, FI-33101 Tampere, Finland
mikko.varila@tut.fi

ABSTRACT

The paper discusses the validity aspect of the ratio type of performance measurement. Quality cost measurement is used as an illustrative example of the challenges that might be faced while striving to maintain validity in practice and possible ways to overcome these challenges. The paper suggests that some surrogate ratios have become so established that their validity has not been seriously considered at the conceptual level. Therefore, they may not be valid measures of the concepts under the primary interest. Threats to validity may be faced at any step from defining the concept and operationalizing the definition to actual measurement, although operationalization was noticed to be the most challenging step. Another important conclusion is that threats to validity seem to arise in ratios if the causal relations between the numerator and denominator are not properly identified or if they are difficult to conceptualize. Validity is a highly context-specific issue and thus the requirements for the causal relationship may depend on the user of the measure.

1 INTRODUCTION

A performance measure can be defined as a metric used to quantify the efficiency and/or effectiveness of an action or an end-result (Neely et al. 1995). The soundness of a measure has typically been evaluated in terms of its validity, reliability, and relevance. A typical performance measure is a ratio which compares outputs with inputs, such as return on investment does. Some ratios have become so established in performance analysis that they are rarely tested for methodological validity (McDonald and Morris 1985). A ratio consists of a numerator and denominator, which are compared with each other. The unit of measurement can be for example money or quantity – depending on whether a phenomenon from a monetary or real process is examined. Uusi-Rauva (1996) has pointed out that companies should place the focus in management on the activities in the real process. However, unambiguous performance measures are typically difficult to construct and in such cases we often settle for surrogates instead of principals (Ijiri 1975, pp. 40-41). This may increase the reliability of measurement if a surrogate can be measured with more accurate and
plentiful data. However, it may also weaken the validity of measurement as one moves further from the concepts in which the primary interest lies.

Quality is a good example of something that is commonly measured as a surrogate ratio derived from the monetary process: the percentage of quality-related costs on sales (Mandal and Shah 2002). The principal is the quality-producing ability of the organization, which ideally would be measured as the amount of quality-related problems in relation to the underlying burden. The surrogate has been used unquestioningly in the research on quality costing (see e.g. Williams et al. 1999) and few alternatives to it have been offered (Mandal and Shah 2002). Nevertheless, this is understandable because it is a rather easily achievable measure and one that gives a good idea of the financial impact of poor quality on business in a given period of time. Even though the surrogate is derived from the monetary process, the main purpose of its use is identifying problems in the real process (Oliver and Qu 1999). However, the connection between the real process and its surrogate measure has not been made clear.

Why then might the monetary and real processes differ from each other to such a degree that they could not be used interchangeably? In the first place, one could question the use of quality costs for correctly representing quality problems, and especially sales for representing the burden. Another problem with surrogate *ratios* is that many cause-and-effect relationships in business come about with a delay. For example, Johnson and Kaplan (1987) have argued that comparing income and expenses in the short run is difficult, which renders the short term measures of profitability invalid. Quality is also a good example of a phenomenon which involves lagged courses of events. Part of the quality failures can be identified immediately after their occurrence, and they are reported as quality costs. Part of the failures go all the way to the customer, in which case identification, reclamation, correction of the problem, and reporting of the costs may be very distant in time from the actual occurrence of a failure. Inability to track all customer needs while developing a product leaves the perception of poor quality to the customer even though no actual failure occurs. Due to such delays, the traditionally used quality costs per sales -ratio is an invalid indicator of what it is intended to measure.

The objective of the study is to discuss the validity aspect of the ratio type of performance measurement. Quality cost measurement is used as an illustrative example. The paper is based on a research project in a single case organization in which a quality costing system was developed in close collaboration with quality and accounting departments. A validity problem was faced while constructing the main-level quality cost measure, which led to the consideration of alternative solutions.

## 2 THEORETICAL BACKGROUND

### 2.1 Validity in performance measurement

Validity is a theoretical concept that refers to the true ability of a measure to illustrate the phenomenon or concept that it is being used to represent (Carmines and Woods 2005). Generally, validity is one of the key attributes that determine the soundness or
appropriateness of a measure. Others include reliability, relevance, and feasibility (see e.g. Uusi-Rauva 1996; Laitinen 1998). Although the concept of validity is often associated with the measure, the instrument of measurement, it really refers to the process of interpretation based on data rather than to the instrument itself. Importantly, validity is not a context-independent attribute of a measure, which reliability is, for instance. Validity should be interpreted in relation to the purpose of measurement (Carmines and Zeller 1981): return on capital may be a valid measure of profitability but is less valid in tracking customer satisfaction.

Essentially portraying the validity attribute, Belkaoui (1993) uses the term “representational faithfulness” instead, which seems like a good and terse definition of the concept. Perfect validity will be achieved if the theoretical and operational definitions of a concept are parallel, but the problem is that – especially concerning abstract concepts – it is extremely difficult to formulate operational definitions that fully represent (are parallel to) the theoretical concept. Consider, for example, the concept of “quality”: absolutely valid measurement of quality would require simultaneous consideration and well-balanced emphasis of a number of different – and as such well justified – interpretations of quality as well as inclusion of the time perspective and the views of separate stakeholders.

Carmines and Zeller (1981) connect validity and nonrandom errors with each other: they argue that invalidity arises because of the presence of nonrandom errors in measurements. Random errors indicate problems with reliability. Obviously, this interpretation is implicitly data-analysis-driven, which means that the validity can only be judged after measurements have been made. Random and nonrandom errors could be identified from the dataset that is already collected. Concerning the interpretation itself, the linkage from systematic or nonrandom error to invalidity seems sustainable. In the case of nonrandom error, the measure is clearly measuring something other than that intended. However, random errors may also be caused by an invalid measure and not just by an unreliable one. Consider, for instance, using the selling price as a measure of product cost. Selling price might be a valid and reliable measure of product market value or something else, but it is clearly an invalid measure of cost. As is well known, the selling price is in many cases – consciously or unconsciously – independent of product cost. That is, companies sell both profitable and unprofitable products. As a result, selling price–based measurement may produce quite randomly both under- and overestimations of product cost. This is seemingly related to invalidity rather than to unreliability of the particular measure.

In considering practical business management and performance measurement needs, discussion on how to identify the possible problems with validity from empirical data is not as important as to understand the issues that drive the validity of measurement more generally. The question then is which issues are potential threats of validity or which issues are likely to improve the validity of measurement.

As validity is both a very conceptual and context-specific phenomenon, it seems quite difficult to give universal guidelines for improving measurement validity. Context-specificity means in practice that a perfectly valid measure may be totally invalid somewhere. Thus, the best that can be said might be “be careful about what and how you infer from the measurement data”. Belkaoui’s (1993) term, representational faithfulness, reminds us that an uncompromising analytical process is needed for
developing good consistency between a theoretical construct and its operationalized definition. However, an uncompromising analytical process is a high-level conceptualization that leaves much room for various practical interpretations. Unfortunately, the literature offers only little discussion regarding threats to validity. McDonald (2005) divides threats into three categories: 1) incomplete definition of the concept, 2) use of an inadequate measure, or 3) unreliable scoring of a measure. While the third category comes quite close to the idea of reliability, the first two are certainly relevant. However, a quite paradoxical dilemma remains: how do we interpret or operationalize “inadequate measure” or “incomplete definition”. For the theory of validity to be normative, operationalization should be possible.

Many financial and non-financial measures – consider e.g. ROI-%, delivery accuracy, total productivity – are actually composed of two or more distinct variables that are proportioned to each other. According to McDonald and Morris (1985), the ratio method can be generally formulated as follows:

$$ \frac{Y_i}{X_i} = a + b_i $$

where $Y_i$ and $X_i$ are the financial variables measured, $a$ is the industry norm, and $b_i$ is the measure of conformity to the industry norm. The above formulation can also be applied to the situation inside a single company by making $a$ the company norm or objective and $b_i$ the measure of conformity to the company norm or objective. Coming back to the issue of validity, utilization of the ratio method means that the problems with the validity of measurement may not only concern distinct variables but also the structure of ratios that are based on two or more variables. The denominator and numerator may include variables that are valid representations of the original concept per se, but when put together, threats to validity may arise. As the meaning of a ratio is often easily understood by intuition, there has not been too much theory development focusing on ratio measurement. As a result, there appears to be inadequate theoretical understanding of the issue. McDonald and Morris (1985) suggest that the theoretical and empirical foundation of ratios is virtually absent. For example, we have found no literature which discusses the possible implications of different types of ratios for measurement techniques and applications: For instance, are the threats to validity different in ratios that contain variables with or without mutual cause-effect links between numerator and denominator.

What is needed is more normative discussion on validity. This paper tries to provide such discussion in one setting – in the context of quality cost measurement.

2.2 Measurement of quality costs

Quality costs have been defined and categorized in many ways. One common typology divides them into the costs of conformance and nonconformance (Crosby 1979). Costs of conformance are costs incurred to make sure a product or service is right the first time. They are needed to implement and maintain a system that aims at eliminating deficiencies, at achieving products with perceived quality, and in that way at ensuring conformance to quality standards. Costs of nonconformance are incurred to correct a problem or irregularity. Thus, they compensate for nonconformance to quality standards. According to a more detailed PAFF model (see e.g. Feigenbaum
the costs of conformance can be divided into prevention costs and appraisal, and the costs of nonconformance into internal and external failure costs. Internal failure costs are discovered before the product is delivered to the customer and external failure costs after that.

Since the introduction of the categories of quality costs, there has been a constant debate concerning the dependencies between the different cost types (see e.g. Ittner 1996; Anderson and Sedatole 1998). Whatever their mutual relationships are, these costs behave very differently as cost objects. Whereas most costs of conformance are inherent in processes designed to deliver something, costs of nonconformance are highly unpredictable and desired to be kept at a minimum. Basically, there must be some costs of conformance in the company to maintain quality. That may be a reason why especially the costs of nonconformance still seem to receive more attention in the research regarding quality-related costs.

Despite the slogan “you get what you measure”, only a rather small proportion of companies report and measure quality costs. Surveys have shown that this proportion has been around 25-50 % of those companies which have responded (Oliver and Qu 1999; Schiffauerova and Thomson 2006). There has been controversy regarding whether quality-related costs are worth measuring, because after all they can never be completely ascertained. However, most researchers have shared the view that measuring quality costs helps in understanding factors affecting quality and in identifying targets for improvement.

The significance of quality costs is easier to understand when they are proportioned to an important figure (Crosby 1979, p. 126). Accordingly, typical quality cost measures are ratios, where total quality costs or part of them are proportioned to a relevant indicator of volume. The common bases of measurement have been reclamation costs, failure costs, warranty costs or total quality costs in proportion to sales, marketing costs, variable costs, manufacturing costs, or total costs (Uusi-Rauva 1996, p. 3; Mandal and Shah 2002). Schiffauerova and Thomson (2006) recommend that a performance measurement system should include both “global” and more detailed metrics. In addition, performance measures may include the costs of preventive quality work and appraisal (Schiffauerova and Thomson 2006). Mandal and Shah (2002) point out that only a few other types of ratios are discussed, such as those of internal failure cost to average production costs, warranty costs to sales turnover, supplier’s appraisal cost to purchased material cost, manufacturing appraisal costs to production cost, and external failure costs to sales turnover. They also criticize most authors for discussing the subject of defining and categorizing quality costs while skipping over the subject of how to actually measure them.

Morse (1983) lists several problems related to reporting of quality costs, including the subjective nature of a large proportion of the information. Possible sources of problems with validity and reliability are caused by the fact that the effort and the achievement are probably not matched in a single reporting period. Instead, there may be a long time delay between manufacturing, assembling, and delivering a product and the perception of its quality costs. In addition, variations in activity may reduce the comparability of quality costs from different periods. This may not be a problem from the financial perspective, but improving quality requires identification of cause-and-effect relationships and thus proportioning quality costs to underlying volume is
necessary. However, because of the aforementioned factors, it is debatable what this volume is and in which period of time it has been achieved.

The aforementioned problems are emphasized in project business where the order-delivery process is typically long, but they are also valid in the case of mass products if the life cycles are long and there is some heterogeneity in production processes. Project business is a challenging environment for quality costing also in the sense that it is difficult to come up with one fixed and measurable parameter, or even a small number of parameters, that could determine the quality of a project. This is because a project as “a temporary endeavor to create a unique product, service or results” (Project Management Institute 2004) is typically more complex than a product and may even include multiple products. Thus, quality costs have often been assessed qualitatively in project environment (see e.g. Malmi et al. 2004), because quantitative data is rarely available from unique projects.

3 CASE STUDY ON QUALITY COST MEASUREMENT

3.1 Current quality cost measurement in the case organization

The research project was carried out within a business unit of a large Finnish company. The business is project-oriented and the projects are carried out both independently and as parts of the parent company’s deliveries. The business unit has the following main functions: R&D, sales, projects, engineering, purchasing, and service. As regards quality costs, the business unit has limited itself to the registering of internal and external failure costs. Internal failures refer also in this case to quality problems perceived by the business unit itself during the execution of a project, which lasts approximately one year. Most of the internal failures are observed in installation, and the cause is often engineering or installation itself. External failures are perceived mainly by the customer during the two-year guarantee period after the project. Problems are most often related to the functioning of a product or a process.

The business unit has a rather advanced quality failure costing system. Each time a quality failure is observed, feedback is given and the failure receives a work number to which the direct costs of fixing the problem are assigned. Feedback is supplied with manifold information, which is intended to provide help in discussing quality costs from as many perspectives as possible. Registered information includes basic information on the project (product line, customer, destination, sales price), product and product family, the activity/department that caused the failure, a description of the failure, and the costs incurred.

![Figure 1: The history of quality costs as percentages of sales.](image-url)
During the research project, the researchers had access to all these data.

The active registration of quality failures and costs has made versatile performance measurement on various levels possible. Quality costs are actively followed from the viewpoints of products, activities, and also individual projects. These measures are important when the quality level of a certain product family or activity is analyzed. They should also help in finding ways to prevent the problems from happening in the future. However, an overall quality cost measure is needed to obtain a general view of the quality-producing ability of the organization. The most actively used measure for this purpose is quality cost as a percentage of sales, divided into internal and external failure costs. This measure is calculated monthly. The development of the measure’s value in recent years is depicted in Figure 1.

3.2 **Validity problem: time-lags between the real and monetary processes**

The managers in the company were dissatisfied with the validity of the current quality cost measure. They had a gut feeling that due to the long lifecycles of a project in their business the quality costs are not comparable with the sales within the same month. Accordingly, explanations for the perceived invalidity were explored first on the basis of the typical lifecycle of an individual project and second on the basis of a time-series analysis at the organizational level.

<table>
<thead>
<tr>
<th>Lifecycle of a project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year -1</strong></td>
</tr>
<tr>
<td>Planning and pre-engineering</td>
</tr>
</tbody>
</table>

**Figure 2: Lifecycle of a typical delivery project.**

The lifecycle of a typical middle-sized delivery project is several years (see Figure 2). A project starts with an offering phase, which often includes planning and pre-engineering, and it may take more than a year before the delivery contract is made. Execution consists of several sub-phases and lasts from less than a year to one and a
half years. From the viewpoint of cash flows, the invoicing of a typical project takes
the form of an "S-curve". Quality problems may occur and they may also be
identified in several phases of a project. The internal failure costs are centered at the
end of, or even a bit after, the execution phase, because the process of dealing with
the problems and fixing them causes delay. External failure costs occur within the
two-year guarantee period after execution. Customers typically make reclamations at
the beginning of the guarantee period when failures are observed, and at the end of it
when they are trying to derive full advantage from the guarantees. However, all
invoicing does not lead equally to quality costs. A small project in terms of invoicing
may cause proportionally considerably more quality costs than a big one, which is
why also the number of projects in hand should be considered. Re-build projects
which replace customer’s old facilities and projects with new products are more
challenging than the others. Failures are more likely to occur when the work load is
high, which can be indicated for example by such measures as the number of
engineering hours and of tasks which are behind time.

When individual projects are piled together, the effects of the aforementioned factors
can be examined at the organizational level. It could be seen visually that a rise in
invoicing, for example, led to a similar rise in quality costs after some delay. In the
case of internal failure costs, this delay was rather short, but regarding external failure
costs the delay was as much as 2-3 years. The same applies also to the number of
projects, the percentage of new products of sales, engineering hours, the number of
tasks late, and somewhat more weakly to other factors. After visual examination the
effects and time lags between variables were considered with statistical tests. A cross-
correlation analysis of the time series was carried out, the purpose of which was to
determine the time lags between the variables as well as the most influential variables.
Table 1 shows the time lags between each variable and internal and external failure
costs where the correlation is greatest. The time lag is expressed as the time in months
between the phenomenon and costs. The results were in line with all the preliminary
assumptions regarding project lifecycle: large volume, difficult projects, new
products, and heavy burden in general will eventually lead to higher quality costs, but
not within the same month.

Table 1. The effect of volume and load variables on quality costs.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Internal failure costs</th>
<th>External failure costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time lag</td>
<td>Correlation</td>
</tr>
<tr>
<td>Invoicing</td>
<td>10</td>
<td>0.678</td>
</tr>
<tr>
<td># of projects</td>
<td>6</td>
<td>0.667</td>
</tr>
<tr>
<td>Engineering hours</td>
<td>13</td>
<td>0.662</td>
</tr>
<tr>
<td>% of new products of sales</td>
<td>16</td>
<td>0.649</td>
</tr>
<tr>
<td># of tasks late</td>
<td>10</td>
<td>0.451</td>
</tr>
<tr>
<td>Internal failure costs</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>External failure costs</td>
<td>-24</td>
<td>0.602</td>
</tr>
</tbody>
</table>

1) Each time series has been calculated as a 6-month moving average.
2) Each correlation is significant with the significance level of 0.001.

In line with the preliminary assumptions of the managers, the validity problem has
emerged with the current quality cost performance measure (see Figure 1). Because of
the long lifecycle of a project (see Figure 2), its quality costs and invoicing are not
likely to be compared within a single month. As a result, the ratio does not proportion
quality costs to a relevant indicator of volume or load. This problem resulted in uncertainty concerning what the measure actually describes. On the one hand, the current quality cost measure is considered useful because it gives a clear impression of the magnitude of the impact that poor quality has on the business. This might be considered as an accounting perspective, because it deals with the actual cash flows in a given month. On the other hand, from the perspective of the quality organization, the current measure gives little information on the actual quality-producing ability of the business unit. For example, a peak in the measure may be a sign of increased quality costs, but also of temporarily decreased sales. However, quality costs today may be a result of failures in projects which were invoiced years ago. From the perspective of quality management, there is no point in this kind of measurement. Thus, alternative measures were developed from the quality improvement perspective.

3.3 Comparing alternative approaches to quality cost measures

Many approaches to solve the timing problem were tried. The first approach is based on the idea that the cash flows from different periods should be moved so that they become comparable. On average, one could say that the current internal failure costs are a result of invoicing that took place X months ago. Likewise, external failure costs would be a result of invoicing that took place Y months ago. Thus, in this approach internal and external failure costs were proportioned separately to invoicing with the time lags obtained from the correlation analysis. In carrying out this study, we found that the average delay between sales and internal failure costs is 10 months. Between sales and external failure costs this was 31 months. Although it is easy to calculate with this measure and to understand it, the results may be distorted because the average time lag does not apply to all projects.

The second approach tries to tackle the problem by taking into account which delivery projects actually may cause internal or external failure costs. As was mentioned in Chapter 3.2, internal failure costs occur during the execution of projects as the problems are noticed by the company itself. Accordingly, the denominator of the measure would be the value (= total invoicing) of the projects under construction. Using the same logic, external failure costs may be caused by the projects under warranty and therefore the total invoicing of these projects would be the denominator in the measure of external failure cost percentage. Thus, the denominator of the measure would be the invoicing of the projects under construction (internal failure costs) or warranty (external failure costs). This approach was successful because it automatically directs attention to the underlying projects and causes of quality problems. The weakness of the method is that it assumes that all the projects under construction or warranty are evenly prone to cause quality costs, which is not the case in reality. There may be projects in very different phases, which is not easy to take into account when calculating the measure.

Thirdly, if it is desired to ensure that quality costs and invoicing are compared from the same projects, a weighted moving average has to be calculated from project-specific quality costs per invoicing rates. In this way one can be sure that quality costs and invoicing are connected with the same phenomena, the same projects. Regarding the practical use of the measure, it is problematic that the project, including its guarantee period, must be completed in order to get the project-specific failure costs
per invoicing rates. As a result, this measure may lag behind and its usefulness may be limited to evaluating the situation in the long run.

![Graph](image.png)

**Figure 3:** Internal (left) and external (right) failure cost percentages calculated with different approaches.

The three alternative approaches were evaluated on the theoretical basis during their construction and in practice during a follow-up period after the research project. In Figure 3 the values of the measures are aggregated separately for both internal and external failure costs. Interestingly, all the approaches can be theoretically rationalized, but yet quite different impressions of quality could be adopted, depending on which measure is used. Based on the figures, the differences are much more significant in the case of external failures than in that of internal failures. Thus, the longer the delay between the costs and the invoicing, the bigger is the effect of alternative measures.

It was clear to the quality department that they could not continue using the traditional approach, because it does not provide them with useful information. Just moving the cash flows with average delays aroused interest, but was not considered reliable enough to be used as an overall measure. Comparing the quality costs with invoicing of the underlying projects received significant approval. It was mentioned that this approach not only appears to be theoretically the most reasoned, but also offers a possibility to direct attention to the underlying projects and drill down to the actual project-specific problems behind the figures. Although the measure does not give dramatically different results, it promotes discussion regarding the past failures and how to prevent them in the future. The project-specific moving average provides this possibility as well, but lagging behind is a significant drawback and may hinder the use of this approach in determining the present-day quality-producing ability.

4 DISCUSSION

On the basis of the conceptual analysis and the findings from the case, validity can be considered in different stages of constructing a performance measure according to Figure 4. Validity can be threatened when one moves from any step to the next step.
The first critical question is whether the concept is defined in a valid way. For example in the case of quality-producing ability, a valid indicator would have to describe the amount of quality (or non-quality) produced in relation to the efforts made. One could question whether it is valid to focus on quality problems or whether the definition should be based on perceived quality, e.g. customer satisfaction. The second step – which our case study especially dealt with – is from conceptual definition to operational definition. The case accepted that quality (failure) costs are sufficiently representative of quality problems. However, finding a comparable load was more challenging. Invoicing can be questioned as representative of load, because the entire load that the invoicing is connected with is not evenly prone to quality problems. Furthermore, causal consistency requires considering the timing of the load, as was discussed in the case part above. Finally, validity may be lost also in the step from operational definition to measuring. For example, a quality costing system which gathers only direct costs of the problems related to delivery projects is not valid in the sense that it systematically misses the indirect costs and the problems in support functions.

1. Concept: An abstract object or thought
2. Definition: A concept put into words
3. Operational definition: Identification rules for a concept’s definition
4. Measuring: Gathering data
Example: Quality-producing ability
Example: The amount of quality problems in relation to the load
Example: Registered quality costs per invoicing during a period
Example: Quality-problem database & accounts receivable

Figure 4. Questions regarding validity in different stages of constructing a performance measure (adapted from McDonald 2005).

The alternatives presented for measuring quality-producing ability represent different stances concerning the causality between the numerator and denominator. Figure 5 depicts three possible interpretations of the relation between two variables in a ratio. First, a ratio may be composed of two variables that are not causally related. For example, the amounts of sales of two distinct periods do not depend on each other. (In real life the nature of the previous period may have some effect on the following one). Second, there might be some causal relations between the numerator and denominator, but these links are weak or fuzzy and sometimes even reciprocal, so it is difficult to separate the cause from the effect. Third, the causality between denominator and numerator is evident and strong, but (as is also the case in other types) there might be some other variables besides the denominator affecting the value of the numerator.
Generally, threats to validity in ratios seem to arise if the causal relations are not properly identified or if they are difficult to conceptualize – and especially if there is a conflict between the concept and the operational definition in the interpretation of causality. In those cases, inferring something based on the measure is likely to be difficult. Coming back to the case material, due to the time lags the traditional operationalized quality cost measure belongs to the category A, although the definition of the concept refers more to category B or even C. On the other hand, the operational definitions and conceptual definitions of three other alternatives seem to be much more in line concerning causality. However, when inferring something on the basis of measurement data of any of these measures, one must bear in mind that there is a number of variables besides those measured in these ratios that certainly affect the amount of warranty costs.

5 CONCLUSIONS

The study discussed the validity aspect of ratio-type performance measurement. An illustrative case on the measurement of quality-producing ability was presented and used for demonstrating both the relevance of the problem and the challenges that are faced in practice while striving to maintain validity.

Quality-producing ability is a good example of a concept that has been conventionally measured as quality costs as a percentage of sales. However, as the case study illustrated, this measure seems not to be valid from the viewpoint of quality management, because the causal relationships assumed by the definition do not hold: due to the project’s lifecycle and delays in cash flows, the ratio does not compare the quality costs and the load from the same project gamut. Alternative approaches were suggested that tackled the validity problem with varying success and with varying practical usefulness. On the basis of the conceptual analysis and the findings from the case, validity was considered in different stages of constructing a performance measure. An especially challenging step in this process was noticed to be operationalizing a definition. That is why we suggest being especially careful regarding this step while constructing ratios.

Another important conclusion is that threats to validity seem to arise in ratios if the causal relations between the numerator and denominator are not properly identified or
if they are difficult to conceptualize. In the case, the measure was not valid from the quality management point of view, because it requires a causal relationship between the quality costs and the load. However, it was more valid from the viewpoint of accounting and finance, which does not require this relationship. This confirms the view that validity is a highly context-specific issue and thus the requirements of the user must be taken into account when constructing a performance measure.

REFERENCES


Crosby, P. B. (1979), Quality is free, New York, McGraw-Hill.


