Journal of Business Research 65 (2012) 1790–1799

Contents lists available at SciVerse ScienceDirect

Journal of Business Research

Multidimensional assessment of organizational performance: Integrating BSC and AHP☆

Alexandre Veronese Bentes [a,b,1], Jorge Carneiro [b,*], Jorge Ferreira da Silva [b,2], Herbert Kimura [c,3]

[a] Praia de Botafogo 528, apt 1108 bloco A Torre E, Botafogo, Rio de Janeiro, RJ, 22.250-040, Brazil
[b] Rua Pacheco Leão, 646 casa 7, Jardim Botânico, Rio de Janeiro, RJ, 22.460-030, Brazil
[c] Rua da Consolação, 930, Consolação, Sao Paulo, SP, 01.302-907, Brazil

ARTICLE INFO

Article history:
Received 1 September 2010
Received in revised form 1 April 2011
Accepted 1 August 2011
Available online 23 November 2011

Keywords:
Performance measurement
BSC
Balanced scorecard
AHP
Analytic hierarchy process
Multi-criteria decision-making

ABSTRACT

Measurement of organizational performance is a complex issue given that performance is a multifaceted phenomenon whose component elements may have distinct managerial priorities and may even be mutually inconsistent. This paper presents the case of a Brazilian telecom company to illustrate and critically analyze the integration of two methodologies, Balanced Scorecard (BSC) – a multiple perspective framework for performance assessment – and Analytic Hierarchy Process (AHP) – a decision-making tool to prioritize multiple performance perspectives and indicators and to generate a unified metric for the ranking of alternatives (in this case, performance of functional units). An iterative and interactive procedure coupled with an agreement-building approach among managers generates priority values for performance dimensions and respective indicators. The paper discusses the advantages and disadvantages of the design.

1. Introduction

Organizations compete for resources and customers and must somehow assess the results of their decisions and actions. Organizational performance is by no means a simple phenomenon; rather, it is a complex and multidimensional concept (Cameron, 1986; Chakravarthy, 1986; Venkatraman & Ramanujam, 1986). Cameron (1986) states that organizational performance is inherently paradoxical because, while a given perspective may indicate good performance, another perspective may indicate the opposite. Oftentimes a given performance indicator can only be improved at the expense of another. Furthermore, individuals may have different preferences about which aspects are most relevant to define and assess performance (Zammuto, 1984) and, as a consequence, may disagree on which measures to use, the level of importance to assign indicators, and how to interpret results.

Given the complexity of the phenomenon, several researchers (e.g., Barney, 2010; Chakravarthy, 1986; Venkatraman & Ramanujam, 1986) advocate the use of multiple perspectives and multiple measures of organizational performance. Kaplan and Norton (1996) advance a framework of performance conceptualization and measurement – Balance Scorecard (BSC) – that explicitly incorporates several dimensions of performance. Although multiple perspectives and measures may better represent the multifaceted nature of the phenomenon, they pose several difficulties: (i) possible need to assign non-equal priorities to perspectives and to performance indicators within each perspective; (ii) need to account for mutually inconsistent results; and (iii) need to design an aggregated metric that would somehow summarize the whole story of success (or failure thereof). One of the methods that can address the complex issues of a balanced system of performance assessment is the Analytic Hierarchy Process (AHP), developed by Saaty (1980, 1990a). AHP relies on decision-makers' knowledge and expressed opinions in order to build a structure of hierarchically-organized objectives, criteria and decision alternatives. Based on decision-makers' explicitly stated priorities across criteria, AHP builds a ranking of the alternatives with respect to the objective and indicates the extent to which each alternative is better (or worse) than each of the others.

The main objective of the present study is to critically illustrate the application in a real-life business setting of an integrated approach that combines two theoretical frameworks – the multidimensional perspective on performance measurement (in this case, BSC) and the multidimensional ranking of decision alternatives (in this case,

☆ The authors thank Angela da Rocha (Pontifical Catholic University of Rio de Janeiro) and JBR anonymous reviewers for their comments and suggestions on earlier versions of this article.

* Corresponding author at: Rua Dezenove de Fevereiro 127 / 201, Botafogo, Rio de Janeiro, RJ, 22.280-030, Brazil. Tel.: +55 21 9163-4177.
E-mail addresses: alexvb@globo.com, abentes@timbrasil.com.br (A.V. Bentes), jorgemtc@iag.puc-rio.br (J. Carneiro), shopshop@iag.puc-rio.br (J.F. da Silva), herbert.kimura@mackenzie.br (H. Kimura).
[1] Tel.: +55 21 8113 1866, +55 21 3507-2320.
[2] Tel.: +55 21 2540 0804.
[3] Tel.: +55 11 2114 8268.

0148-2963/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbusres.2011.10.039

AHP) – in order to comparatively assess the performance of functional subunits within an organization. A subsidiary objective is to extend the external validity of both BSC and AHP by applying those frameworks to a particular industry (telecommunications) and setting (Brazilian environment and managers), which, thus far, the literature has rarely examined. Thus the paper focuses on the practical implementation of an integrated framework, bringing together two well established mechanisms for performance analysis and decision making. Although other studies explore a joint application of AHP and BSC, they typically rely on illustrative hypothetical examples (e.g., Leung, Lam, & Cao, 2006) or on a theoretical discussion of potential uses (e.g., Jovanovic & Krivokapic, 2008). In contrast, this paper examines a real-case implementation of AHP-BSC. Although one can conceive of several levels of organizational performance (e.g., corporate, business unit, functional area, etc.), the present study assesses only the performance of three functional areas in the financial department of a Brazilian privately-held telecommunications company.

The contribution of this paper rests on the attempt to address the thorny issue (*cf.* Venkatraman & Ramanujam, 1986) of organizational performance measurement — in particular, how to (i) make sense of multiple (and often mutually conflicting) perspectives and measures of the phenomenon; (ii) prioritize among them; (iii) reach some synthesized assessment; and (iv) rank-order the level of performance of alternatives under comparative evaluation. A real-life example illustrates the procedure based on agreement-building among managers, which has the additional benefit of enlightening their knowledge as to performance outcomes, thereby developing a shared vision, and fostering a sense of *esprit de corps*. The study of a Latin American firm is relevant because, as argued by Brenes, Haar, and Requena (2009), the process of internationalization and openness of economies in this region has far-reaching impacts on management practices. In this context, formal structures related to strategy formulation and implementation emerge as fundamental elements for Latin American firms since the mid-1990s (Brenes et al., 2009). When the internal political and economic environment becomes more stable, as in the case of Brazil, formal policies and decision-making tools become more useful. In particular, the analysis of a real application of an integrated BSC and AHP framework to assess performance exemplifies the pattern of progress in Brazilian management practices.

The following sections demonstrate the rank-ordering of the overall performance of the functional areas of the financial department of this Brazilian telecommunications company, taking into consideration (i) the relative degree of importance of four distinct perspectives of organizational performance; (ii) the relative degree of importance of performance indicators within each perspective; and (iii) the comparative relative performance of three functional areas in light of a weighted combination of all performance indicators and dimensions.

2. Theoretical framework

The theoretical framework comprises three parts: (i) organizational performance in general and the balanced scorecard (BSC) in particular; (ii) decision-making with multiple and often mutually inconsistent criteria in general; and (iii) the analytic hierarchy process (AHP) as a particular tool for multiple criteria decision-making.

Organizational performance is a complex and multidimensional phenomenon. Several scholars (e.g., Venkatraman & Ramanujam, 1986) argue that traditional financial measures are insufficient to assess properly the performance of organizations and to provide guidance for strategic action (Kaplan & Norton, 1996). Financial measures usually employed by firms – for example, several measures of return (ROA, ROI, ROE) or margins (gross margin, net margin, ROS) – indicate past results only and say little about expectations of future performance. Some financial measures, however – such as NPV (net present value), Tobin's *q* or share value – do reflect expected long-term future performance (Barney, 2010).

Ambler (2000) contends that performance is a function of both external (e.g., profit and financial measures adjusted by the change in brand equity) and internal (e.g., innovation health and employee commitment) market metrics. Regarding the related concept of export performance, Zou, Taylor, and Osland (2008) argue for a conceptualization based on financial, strategic and satisfaction-based views. In an application of AHP to assess export performance, Diamantopoulos and Kakkos (2007) posit that a hierarchy of four levels is adequate to structure performance assessment: (i) the overall goal (i.e., export performance); (ii) different (export) objectives (e.g., sales versus profit versus new product introduction); (iii) the frame of reference (e.g., own plan versus competition versus customers); and (iv) time frame (i.e., short-term versus long-term).

Also, if a construct seems to entail two or more facets (e.g., performance), the researcher must determine, based on a substantive sense, whether a reflective or a formative perspective appears more plausible (Bollen & Lennox, 1991; Diamantopoulos, 1999). A reflective measurement perspective assumes that the (observed) items are the effects of an underlying latent construct; in contrast, a formative measurement perspective assumes that the items "cause" a latent construct. Also under a formative perspective, a change in the value of one of the indicators does not necessarily associate with a change in all other indicators. More detail follows ahead, but for the time being, AHP implicitly assumes a formative perspective of measurement, since the value of the overall or unified metric (e.g., functional unit performance) comprises the contribution of each and every value of the lower nodes (i.e., the performance indicators of each of the four perspectives of the BSC). As such, a change in a performance indicator leads to a change in the synthesized metric, but not vice versa.

Neely (1999, p. 222) contends that, although "it is widely accepted that business performance is a multi-faceted concept […], it is not obvious which measures a firm should adopt" or how measures change over time. For the choice of measures, although several approaches derive from corporate strategy (e.g., Kaplan & Norton, 1996; Keegan, Eiler, & Jones, 1989; Neely, Gregory, & Platts, 1995), Neely (1999) contends that the implementation of measurement systems and their application to managing business performance require further research. Kaplan and Norton (1992) propose that organizational performance be simultaneously assessed from distinct, albeit complementary, perspectives (originally defined as Financial, Customer, Internal Business Processes, and Innovation and Learning), comprising the so-called Balanced Scorecard (BSC) framework. Kaplan and Norton (1992) argue that BSC provides a balance between (i) short-term and long-term objectives, (ii) financial and non-financial measures, (iii) lagging and leading indicators, and (iv) internal and external performance perspectives. However, the very contribution of the BSC – that is, the explicit consideration of multiple performance perspectives other than just a strictly financial standpoint – also brings complexity to the measurement of performance, especially in terms of information overload, judgment biases, and the need to reach some synthetic judgment that summarizes and makes sense of BSC's multiple perspectives and indicators (Chan, 2006). Multi-criteria decision-making frameworks are appropriate for tackling this complication to performance evaluation and decision-making.

Certain decision-making problems involve choices among alternatives that are comparable using multiple criteria, and as such, may present some degree of mutual inconsistency. For example, the values achieved with certain criteria may suggest a decision that might be different from one based on the values of other criteria. In addition, improvement of results under a given criterion sometimes only occurs at the expense of another. This picture, apparently internally inconsistent, is inherent to the assessment of the multifaceted

phenomenon of organizational performance and poses difficulties for the choice – of the best performing unit for instance – since a given business unit is not usually superior in all relevant criteria. In fact, many decisions, including performance assessment, depend not on the simple maximization of a single objective function (criterion), but rather, on some balance among multiple objective functions.

Multi-criteria decision-making tools comprise two general research streams: the American School and the European School. The American School requires that all alternatives be comparable and rank-ordered, and forces decision-makers to reach an explicitly hierarchical structure among the alternatives. In contrast, the European School accepts that alternatives may not always be comparable and that decision-makers may not explicitly rank priorities or even determine a best decision. Well-known decision-making tools of the American School include AHP (Analytic Hierarchy Process; Saaty, 1980, 1990a) and MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique; Bana e Costa & Vansnick, 1994). Important European School tools include ELECTRE (*Elimination et Choix Traduisant la Réalité*; Roy, 1968) and PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations; Roy & Vincke, 1984; Vincke, 1992).

In particular, AHP is useful in prioritizing decision alternatives (Lu, Madu, Kuei, & Winokur, 1994) and may be the most widely used technique for multi-criteria decision-making (Madu, Kuei, & Madu, 1991). According to Liberatore, Monahan, and Stout (1992), AHP is effective in addressing many types of problems that involve multiple criteria, including comparative analysis of performance in business sub-units (Chan & Lynn, 1991), which is the main objective of the present study. As to the scope of application of AHP, Forman and Gass (2001) describe several cases of application of the AHP in organizations (e.g., Xerox, Nasa, IBM, Air Products, etc.). Saaty (1980) argues that five options (equal, weak, strong, very strong, and absolute) represent managers' ability to make qualitative distinctions; furthermore, a compromise between every two successive options enables the use of a 9-point scale. Deployment of AHP in real-life decision making involves successive comparisons between every two alternatives, criterion by criterion, according to a 9-point scale as presented in Fig. 1. If an alternative $\mathbf{A}_i$ is preferable to an alternative $\mathbf{A}_j$, then the value of the comparison scale $\mathbf{P}_c(\mathbf{A}_i,\mathbf{A}_j) = \mathbf{a}_{1j}$ indicates the intensity of relative importance of $\mathbf{A}_i$ over $\mathbf{A}_j$ assigned by the decision maker. The scale allows, in a pairwise comparison, the investigator to establish which alternative is better. Higher values of $\mathbf{a}_{ij}$ indicate stronger preference of alternative $\mathbf{A}_i$ over $\mathbf{A}_j$. The comparison of one pair of alternatives at a time for each decision criterion at a time (instead of a simultaneous comparison involving all alternatives and criteria) reflects the assumption that a decision maker can more easily reveal the preferred alternative by analyzing one property of one object at a time (Voronin, 2007).

By considering all possible pairwise comparisons between alternatives, a matrix **A** results that can represent the relative importance $\mathbf{a}_{ij}$ of each element over each other. Given that an element is as important as itself and taking into account the theorem of reciprocity, then $\mathbf{a}_{ij} = 1$, if $i = j$ and $\mathbf{a}_{ij} = \frac{1}{\mathbf{a}_{ji}}$, if $i \neq j$. The calculation of weights relies on an iterative process in which matrix A is successively multiplied by itself, resulting in normalized weights, w_i. The process halts when the difference of weights between successive iterations is smaller than a given halt criterion. In this framework, w_i represents the importance of alternative A_i relative to all other alternatives. Note that the normalization process entails that weight components total 1.0.

The judgment of decision makers in pairwise comparisons may present inconsistencies when taking into consideration all alternatives simultaneously. In order for the comparison matrix to be consistent, $\mathbf{a}_{ik} = \mathbf{a}_{ij}.\mathbf{a}_{jk}$. However, decision-makers, when comparing alternatives, often violate this relationship. The consistency index (CI) and the consistency ratio (CR) (Sharma & Bhagwat, 2007) measure the degree to which judgments are not coherent. If CR < 10%, then the degree of consistency is satisfactory (Saaty, 1990b).

As a framework of analysis, AHP comprises three main parts: (i) decomposition, (ii) comparative judgments, and (iii) synthesis. Decomposition consists of presenting a complex problem in terms of simpler elements, that is, the overall objective, the alternatives from which to select, and the criteria (and respective sub-criteria) that contribute to reaching the overall objective. Then, the various criteria undergo a series of pairwise comparisons whereby each child item below a given parent node receives a score that represents its relative importance to another child item in the same parent node. The highest-level node constitutes the overall objective. These comparative judgments are actually ratios of degrees of importance. At each parent node, the components of the normalized eigenvector are local weights of the underlying criteria below that parent node and represent the relative and normalized contribution of each to the respective parent node. Thus, relative performance under each criterion is a function of the comparison of each decision alternative. Synthesis of all pairwise comparisons among alternatives, weighed by the importance of respective criteria, helps make sense of how much each of the decision alternatives contributes (in a relative manner) to the attainment of the overall objective. A single metric that synthesizes the multiple judgments (often not mutually consistent) results from calculations of the ratios.

Implicitly employing a formative perspective of measurement, AHP emerges as a useful mechanism for implementation and management of performance systems, particularly the BSC. AHP allows for (i) the consideration of multiple metrics, taking into account a hierarchical analysis of various indicators within various perspectives, and (ii) the establishment of a final composite to compare

Intensity of relative importance	Definition	Explanation
1	Equal importance	The two alternatives contribute equally to the objective
3	Moderate importance of one over the other	Experience and judgment slightly favor one over the other
5	Essential or strong importance	Experience and judgment strongly favor one over the other
7	Demonstrated importance	One is judged much more important than the other
9	Extreme importance	The evidence favoring one over the other is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values	When some compromise is needed

Source: Saaty (1986)

Fig. 1. Scale of comparisons of the AHP.

performance of business units. Despite the relative acceptance of AHP in academic and managerial communities, Smith and von Winterfeldt (2004) and Chan (2006) note that AHP suffers severe criticism on several fronts: (i) lack of axiomatic foundation (though see Saaty (1986) for formalized arguments based on theorems and lemmas of measure theory); (ii) ambiguity of the questions used to elicit answers from decision-makers; (iii) imprecision of the scale used to rate relative preference; and (iv) possibility of rank reversal (Goodwin & Wright, 2000; Harker & Vargas, 1987) if the researcher adds new alternatives to the analysis. Some of the assumptions that AHP relies upon are the object of intense debate, pro and con, as exemplified by conflicting studies (e.g., Bernhard & Canada, 1990; Dyer, 1990; Forman & Gass, 2001; Harker & Vargas, 1987; Lane & Verdini, 1989; Leskinen, 2000; Saaty, 1986, 1999; Schoner & Wedley, 1989). Notably, such discussions among experts afford enhancements to AHP as well as developments to multi-criteria decision theory.

AHP is useful in a BSC analysis. BSC, while considering several relevant dimensions of organizational performance, does not formally explain how to weight their importance in a comprehensive framework. Leung et al. (2006) point to the relative abundance of studies regarding the structure of BSC compared to the dearth of research on protocols for proper implementation. As Reisinger, Cravens, and Tell (2003) p. 431 observe, "Without any information to the contrary, managers might likely assume that since the measures are related and have the same primary objective, then they should be equally important." In practice, however, perspectives and indicators seldom have equal importance. AHP, as a useful tool to prioritize and consolidate performance metrics based on multiple criteria, can be a promising mechanism to help overcome the limitations of BSC.

Some studies do present joint applications of AHP with BSC. Jovanovic and Krivokapic (2008), for instance, use AHP to identify key performance indicators of the perspectives of BSC. Leung et al. (2006) structure a BSC framework with the aid of AHP, considering traditional problems in implementation such as the dependence relationships among attributes and the need to use objective and subjective measures. Applications of AHP within a BSC framework also appear (i) in Chan (2006), who discusses a case of hospital performance assessment; (ii) in Varma, Wadhwa, and Deshmukh (2008), who apply these tools to evaluate performance of the supply chain of oil companies; and (iii) in Sharma and Bhagwat (2007), who describe the assessment of five supply chains in different sectors.

3. Methods and data

The present study investigates the relative performance of three functional units of the financial department of a Brazilian telecommunications company: the Fraud unit, the Collection unit, and the Revenue Assurance unit. Managers consider these units to be the most strategic areas of the financial department and the most influential in the performance of the firm. At the time of the study, one of the authors of this paper was the manager of the Collection unit and was the researcher who collected the data, as detailed below. In order to select the performance indicators and conduct the comparative judgments, the researchers invited sixteen employees to participate in the study: three senior managers (one for each unit, all with long experience in the company); three supervisors (one for each unit); nine senior analysts (three for each unit); and the financial operations director. The diversity in hierarchical levels was important in order to reach a more consensual and department-wide opinion about performance that could better represent the diversity of views in the financial department.

Forman and Peniwati (1998) discuss some possible paths for constructing the comparison matrices, that is, for aggregating individual responses in terms of a group response. Two of these paths are (i) simple averaging across the assessments (of the values of priorities and of the performance level of each alternative in each indicator) produced independently by the evaluating judges, and (ii) an agreement-building approach whereby evaluating judges reach some (pseudo-)consensus about the value of priorities and of performance levels. While some researchers employ the averaging approach (e.g., Chou, Lee, & Chung, 2004; Javalgi, Armacost, & Hosseini, 1989), others prefer the agreement-building approach (e.g., Fletcher & Smith, 2004; Shahin & Mahbod, 2007; and Kumar & Bhagwat, 2007, who employ the opinions of the majority of the interviewees). The present study uses the agreement building approach, which resembles a focus group design. Note that the focus group approach may lead to a pseudo-agreement, since the dynamics of face-to-face meetings may influence individual judgment. On the other hand, Morgan and Krueger (1993) suggest that the interaction among individuals in a focus group is advantageous, since the group effect may generate valuable data deriving from both the consensus and diversity of the participants.

In the first semester of 2009, five meetings with the managers took place. In the first meeting, in accordance with instructions, the sixteen executives and employees selected which performance indicators would serve as the basis for the subsequent analysis. The multifaceted nature of performance assessment calls for an approach that takes into account diverse perspectives. The study used BSC due to widespread use of the technique in outlining the different dimensions of performance. Thus, Financial, Customer, Internal Business Processes, and Learning and Innovation were perspectives included a priori and interviewees did not question them. In preparation, the investigators debriefed interviewees on the BSC system, the AHP methodology, and the research objectives. After the meeting, interviewees received support material on the BSC approach to performance measurement and the AHP methodology of multi-criteria decision-making. Ten days later, all participants met with one of the researchers, who took questions and provided explanations concerning BSC and AHP. The financial operations director then explained the company objectives for the year, with particular emphasis on the financial department objectives. All participants were free to suggest indicators for each of the four BSC perspectives. In order to allow for a comprehensive preliminary analysis and diversity of variables, participants proposed indicators based on the support materials and on their own experience and opinions. Participants left their suggestions anonymously over the course of the days following the meeting in a physical mailbox — such a precaution (anonymity) was important in order for participants not to feel afraid or ashamed, and to avoid attributing (inadvertently or otherwise) more weight to the opinions of hierarchically superior participants. The free suggestion of indicators helped tailor to the context of the firm the variables proposed by Kaplan and Norton (1996), Lipe and Salterio (2000), Banker, Chang, and Pizzini (2004)), Akkermans and van Oorschot (2005), among others.

In the third meeting, a brainstorm session, all of the sixteen participants suggested indicators and posted them on a board. The researchers addressed issues concerning the adequacy of indicators and their classification within perspectives. Participants discussed which indicators appeared key to reflecting the objectives of the financial department, and each participant selected at least two indicators for each of the four BSC perspectives. In the fourth meeting, the group fine-tuned the performance indicators and discussed goals. Finally, in the fifth meeting, the group reached a consensus as to which performance indicators to use. Due to the quantity and diversity of indicators, an intense debate ensued. The discussion eventually arrived at a set of variables that reflects specific characteristics of the firm and its environment, especially from the Internal Business Processes perspective. The final version of the BSC included a selection of nine indicators. Despite the fact that these nine indicators may not fully encompass the complexity, they are reasonably representative of the phenomenon.

The researches then collected two types of data: (i) the values for the nine performance indicators selected for the study; (ii) the relative importance of each indicator versus the others, and of each BSC perspective versus the others. Internal company reports provided performance indicator data. To establish relative importance of the indicators, the participants met five additional times to reach a consensus as to the pairwise comparisons of perspectives and indicators. Importantly, an indirect benefit of the study is a more comprehensive knowledge of performance measurement on the part of managers, a learning experience that emerged from the intense debate. In order to allow participants to express their opinions more freely, the top financial officer did not participate in the final round of meetings.

Note that the definition of performance measures in a BSC mechanism is an open issue. Even a thorough discussion to obtain consensus may not generate adequate indicators. For instance, common measure bias may preclude the inclusion of measures that are unique to a given business unit (Libby, Salterio, & Webb, 2004, Lipe & Salterio, 2000). As an analytical method for deriving weights in a multidimensional setting, AHP has special applicability to BSC. However, the literature also addresses other mechanisms to introduce weights in BSC. For instance, Ittner, Larcker, and Meyer (2003) describe a subjective weighting process in a BSC where managers have some discretionary power to define bonus awards; DeBusk, Brown, and Killough (2003) discuss a hypothetical dashboard case presented to graduate students in which, through principal components analysis, researchers can estimate relevant perspectives and their weights.

Unlike the real-life situation of this study, some works rely on hypothetical cases to elicit weights. Lipe and Salterio (2000) and Banker et al. (2004) ask participants to assume the roles of senior executives, evaluate performance of business units along multiple indicators, and make recommendations concerning manager promotions. In this "laboratory" setting, they found that respondents attribute more weight to common measures (indicators that are similar across units) relative to unique measures (indicators specific to a given unit).

Some aspects of the method deserve discussion. For example, the assessment process occurs on a relative basis, so that even if all areas improve, some will still (relatively speaking) underperform; so, the research should conduct both a relative and an absolute assessment. Also, the present methodology does not explicitly consider a discussion of possible thresholds (minimum or maximum) beyond which the impact of a given performance indicator on overall assessment of performance stabilizes; that is, overall assessment may not be a linear or even monotonic function of individual performance indicators. Yet, managers' subjective attributions of the relative achievement of each unit may implicitly overcome this limitation. Too, the methodology implicitly assumes that any loss in a given performance indicator can be compensated for by a gain of equal magnitude (or weighted compensation) in another indicator. The framework also implicitly assumes that criteria are mutually independent, that is, interactions among them are absent. In other words, the researcher can compare two indicators regardless of the value of other indicators. Note also that in the present study, one of the researchers was also the manager of one of the organizational units and, admittedly, that unit (Collection) earned the highest ranking. In this sense, the researcher might have inadvertently biased discussions.

Since this study focuses on three functional units of the financial department of a Brazilian telecommunications company, immediate generalization of the final results to other departments or to other firms is not possible. The objective of the present study, however, is to evaluate the applicability of a combined framework – BSC and AHP – to business performance assessment, illustrated with a specific real-life example. Therefore, from an academic perspective, the procedure is more important than the results per se. For managers of the company under investigation, however, the results are of course relevant and useful.

4. Application of the analytic hierarchy process

Application of AHP to rank-order the three functional units required nine steps. In *Step 1*, the sixteen participants selected nine performance indicators (Fig. 2) over the course of a series of five meetings. Taken jointly, the indicators provided fair and balanced coverage of the four perspectives of BSC. The AHP hierarchical structure for this study appears in Fig. 3.

In *Step 2*, a necessary result was to reach agreement among all participants regarding the degree of importance of each performance indicator vs. each of the others in the same BSC perspective (parent node). Instead of having participants define the weights individually, the research required that participants jointly discussed why a given indicator would be more or less important than another and the degree of the difference. According to Saaty (1989), the AHP, as a group decision support tool, promotes effective group interaction and participation. By using an agreement-building approach, weights may reflect a more balanced perception of the relative importance of indicators and perspectives. The discussion among managers and analysts had the additional benefit of forcing each participant to provide an explicit justification for his/her reasoning, and also made participants aware of aspects (as commented by other participants) they might otherwise have overlooked had the open discussion not occurred.

In *Step 3*, researchers calculated the local normalized weights for each performance indicator. Table 1 presents the final matrix of performance indicators (sub-criteria) and respective normalized weights. In the 3×3 matrix of the financial perspective, the consistency ratio (CR) is below the threshold of .10, which is acceptable (Saaty,

BSC Perspective	Indicator	Definition	Measurement unit
Financial	A	Recovered value or avoided loss	Reais (Brazilian currency)
	B	ROI of implemented projects	% (in the respective year)
	C	Decrease in operating costs	Reais (Brazilian currency)
Customer	D	Satisfaction level of internal clients	Three-point semantic-differential scale
	E	Satisfaction level of external clients	Three-point semantic-differential scale
Internal Business Processes	F	Number of improvement projects implemented in the year	Cardinal number
	G	Number of new projects fully delivered in the year	Cardinal number
Innovation and Learning	H	Employee skill development	Number of employees that obtained a graduate degree
	I	Training and skill development activities	Number of courses attended in the year by the employees

Fig. 2. Performance indicators selected for the present study.

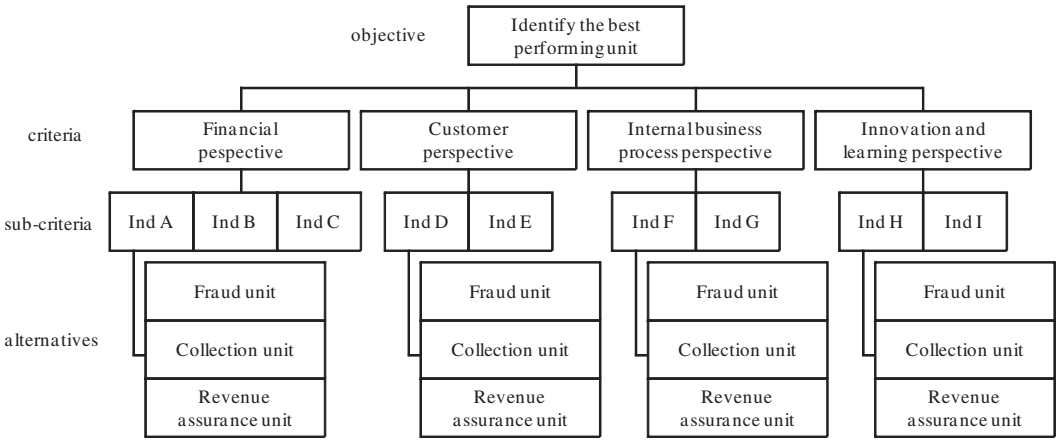

Fig. 3. AHP structure for the present study.

1990b). The remaining three matrices are all 2×2 and (by definition) their consistency ratios are zero.

Interestingly, after the mathematical manipulation necessary to arrive at the eigenvector of each matrix, indicator A (recovered value or avoided loss) ranks as almost 3.5 times (i.e., .71/.21) more important to the financial perspective than indicator B (ROI of implemented projects) and about nine times more important than indicator C (decrease in operating costs). In the Customer perspective, indicator E (satisfaction level of external clients) ranks as approximately seven times more important than indicator D (satisfaction level of internal clients). In other words, the company pays much more attention to external versus internal customers — a reasonable posture given its status of privately-held for-profit firm. As for the internal business processes perspective, indicator F (number of improvement projects implemented in the year) ranks as three times more important than indicator G (number of new projects fully delivered in the year). Based on the company's current situation, emphasis on improvement projects appears reasonable since the finance department is of a maturity level that does not require new and costly projects. In contrast, enhancing existing projects may add value to the firm. Finally, for the innovation and learning perspective, indicator H (employee skill development) ranks almost five times as important as indicator I (training and skill development activities). The results suggest that the financial operations team is already capable of handling its tasks without the need for training programs. Indeed, turnover is low, and the team comprises experienced employees who would profit from further development of their current expertise.

In *Step 4*, managers defined the relative weights of each BSC perspective (vs. each of the others) using the 9-point scale; in *Step 5*, researchers calculated the normalized weights of each perspective in a similar fashion to Step 3. In *Step 6*, a check of the consistency ratio of each comparison matrix indicated whether managers should participate in a new round of preference definitions (paired judgment) to reduce inconsistencies. In a few cases, the consistency ratio was above 10% and participants reconvened to reassess pairwise judgments. Table 2 shows raw and normalized weights, as well as the consistency ratio of the 4×4 matrix. Note that managers judged the financial perspective as much more important than any of the other performance perspectives. In fact, the financial perspective ranks as almost three times (i.e., .59/.21) as important as the customer perspective, about five times as important as the internal business processes perspective, and over seven times as important as the innovation and learning perspective.

All matrices provide examples of the incongruence present in the managers' judgments. For example, in the first line of the matrix in Table 2, the innovation and learning perspective appears to be 4/5 as important as the processes perspective. However, in the last line of the same matrix, the innovation and learning perspective garners half the attributed importance as that attributed to the processes perspective. Rather than representing a weakness of the AHP method, the acceptance of incongruence across judgments – a fact of managerial cognitive limitations and decision heuristics – and the explicit estimation of the magnitude of the incongruence are advantages of the method. If the consistency ratio is too large (i.e., consistency too low), then decision-makers must come back to the drawing board and

Table 1
Relative importance, normalized weights and consistency ratios at the performance indicators (sub-criteria) level.

Financial	Ind A	Ind B	Ind C	w	Processes	Ind F	Ind G	w
Ind A	1	4	7	0.71	Ind F	1	3	0.75
Ind B	1/4	1	3	0.21	Ind G	1/3	1	0.25
Ind C	1/7	1/3	1	0.08	CR	0		
CR	0.03							

Customer	Ind D	Ind E	w	Innovation	Ind H	Ind I	w
Ind D	1	7	0.12	Ind H	1	5	0.83
Ind E	1/7	1	0.88	Ind I	1/5	1	0.17
CR	0			CR	0		

Table 2
Weights and consistency ratio at the BSC perspectives (criteria) level.

Perspective	Finance	Customer	Processes	Innovation	w
Finance	1	5	4	5	0.59
Customer	1/5	1	3	3	0.21
Processes	1/4	1/3	1	2	0.12
Innovation	1/5	1/3	1/2	1	0.08
CR	0.09				

Table 3
From local weights to global weights of each indicator.

Perspective	Indicator	Weight		Perspective	Indicator	Weight	
Finance		Local	Global	Processes		Local	Global
Weight	A	0.71	0.42	Weight	F	0.75	0.09
0.59	B	0.21	0.12	0.12	G	0.25	0.03
	C	0.08	0.05				

Perspective	Indicator	Weight		Perspective	Indicator	Weight	
Customer		Local	Global	Innovation		Local	Global
Weight	D	0.12	0.03	Weight	H	0.83	0.07
0.21	E	0.88	0.18	0.08	I	0.17	0.01

Table 4
Relative degree of success in performance indicators by each functional unit.

Ind A	Fraud	Collec.	Rev. As.	w
Fraud	1	1/7	1/2	0.09
Collection	7	1	5	0.74
Revenue assurance	2	1/5	1	0.17
CR	0.01			

Ind B	Fraud	Collec.	Rev. As.	w
Fraud	1	4	3	0.61
Collection	1/4	1	1/3	0.12
Revenue assurance	1/3	3	1	0.27
CR	0.06			

Ind C	Fraud	Collec.	Rev. As.	w
Fraud	1	1/4	1/2	0.13
Collection	4	1	4	0.66
Revenue assurance	2	1/4	1	0.21
CR	0.05			

Ind D	Fraud	Collec.	Rev. As.	w
Fraud	1	1/5	3	0.19
Collection	5	1	7	0.73
Revenue assurance	1/3	1/7	1	0.08
CR	0.06			

Ind E	Fraud	Collec.	Rev. As.	w
Fraud	1	1/4	2	0.22
Collection	4	1	3	0.63
Revenue assurance	1/2	1/3	1	0.15
CR	0.09			

Ind F	Fraud	Collec.	Rev. As.	w
Fraud	1	1/2	1/5	0.12
Collection	2	1	1/4	0.20
Revenue assurance	5	4	1	0.68
CR	0.02			

Ind G	Fraud	Collec.	Rev. As.	w
Fraud	1	1/2	1/6	0.11
Collection	2	1	1/3	0.22
Revenue assurance	6	3	1	0.67
CR	0			

Ind H	Fraud	Collec.	Rev. As.	w
Fraud	1	1/5	1/3	0.10
Collection	5	1	3	0.64
Revenue assurance	3	1/3	1	0.26
CR	0.03			

Ind I	Fraud	Collec.	Rev. As.	w
Fraud	1	1/4	1	0.18
Collection	4	1	2	0.58
Revenue assurance	1	1/2	1	0.23
CR	0.05			

reassess the relative importance of each criterion until they arrive at some reasonably low consistency ratio.

In *Step 7*, the global weight of each performance indicator – that is, its contribution to the overall objective – is the product of performance indicator local weight times the respective BSC perspective local weight (see Table 3). For example, the global weight of indicator G is .25*.12 = .03.

Step 8 involved the paired comparisons of the performance of each functional area (alternatives in the AHP hierarchy) in terms of each performance indicator. Company reports provided secondary data on the nine indicators for all three functional areas. Although such data is useful for the definition of the relative performance among areas, some degree of subjective judgment (characteristic of the AHP) is certainly involved — since the original scales for producing indicator measurements needed conversion to the AHP 9-point scale. Table 4 presents the relative evaluation of each functional unit (raw and normalized weights) in each of the nine performance indicators, as well as consistency ratios of each 3×3 matrix.

Regarding the financial perspective, the Collection unit ranks as appreciably more successful than the other two units in terms of indicator A (recovered value or avoided loss) and C (decrease in operating costs). The Fraud unit, in contrast, ranks as considerably more successful in indicator B (ROI of implement projects), but the least successful in indicators A and C. In regard to the customer perspective, the Collection unit also performs much better than the other two units, both in indicator D (satisfaction level of internal clients) and indicator E (satisfaction level of external clients). The Revenue Assurance unit performs the worst in both indicators. With respect to the internal business process perspective, the Revenue Assurance unit ranks considerably better than the other two in terms of indicator F (number of improvement projects implemented in the year) and indicator G (number of new projects fully delivered in the year). The Fraud unit performs the worst in both indicators. In terms of the innovation and learning perspective, the Collection unit again ranks more highly than the other two units, in both indicator H (employee skill development) and indicator I (training and skill development activities). The Fraud unit ranks worse (though not by much) than the Revenue Assurance unit.

As the previous discussion shows, looking only at the performance of units (functional areas) in terms of specific indicators, it is hard to tell which unit is best. Results are mixed, since a unit can rank well in a given indicator but not in another. Crucially, AHP allows for the identification of an overall best alternative, in this case, a unit of the financial operations department, the best performer when the researcher considers simultaneously the relative global importance of all indicators. Thus, after ascertaining how well each functional unit (the alternatives in the AHP hierarchy) did in each performance indicator (sub-criteria in the AHP structure) – as shown in Table 5 – the final step (*Step 9*) was to determine the extent to which each

Table 5
Partial contribution of each functional unit to the overall performance objective.

Finance	Ind A	Ind B	Ind C		Indicator			Results	
Fraud	0.09	0.61	0.13		Ind A	0.42		Fraud	0.13
Collection	0.74	0.12	0.66	×	Ind B	0.12	=	Collection	0.35
Revenue assurance	0.17	0.27	0.21		Ind C	0.05		Revenue assurance	0.11
								w	0.59

Customer	Ind D	Ind E		Indicator			Results	
Fraud	0.19	0.22		Ind D	0.03		Fraud	0.05
Collection	0.73	0.63	×	IndE	0.18	=	Collection	0.14
Revenue assurance	0.08	0.15					Revenue assurance	0.02
							w	0.21

Processes	Ind F	Ind G		Indicator			Results	
Fraud	0.12	0.11		Ind F	0.09		Fraud	0.01
Collection	0.20	0.22	×	Ind G	0.03	=	Collection	0.03
Revenue assurance	0.68	0.67					Revenue assurance	0.08
							w	0.12

Innovation and learning	Ind H	Ind I		Indicator			Results	
Fraud	0.10	0.18		Ind H	0.07		Fraud	0.01
Collection	0.64	0.58	×	Ind I	0.01	=	Collection	0.05
Revenue assurance	0.26	0.23					Revenue assurance	0.02
							w	0.08

Table 6
Final results for assessment of the best performing functional unit.

Unit	Finance	Customer	Process	Innovation	Total
Fraud	0.13	0.05	0.01	0.01	0.20
Collection	0.35	0.14	0.03	0.05	0.57
Revenue assurance	0.11	0.02	0.08	0.02	0.23
Total	0.59	0.21	0.12	0.08	1.00

functional unit contributed to each BSC perspective (criteria in the AHP structure) and then how each performs on a global basis (the overall objective in the AHP structure). In *Step 9*, the best performing unit in each perspective derives from a comparison of the vector product between the local weights of the indicators of the business unit (i.e., the achievement of each alternative in each lower-level sub-criteria) and the respective global weights of the indicators. The larger this product, the better the overall performance of the functional unit.

Table 5 indicates the contribution of each unit to the weight of each perspective: e.g., the Fraud unit contributes with .13, the Collection unit with .35, and the Revenue Assurance unit with .11, which totals .59, which is the importance of the financial perspective in the overall analysis.

Similarly derived was the contribution of each unit to the final analysis (i.e., to identify the best performing unit, which is the overall objective). Table 6 summarizes the results.

Results suggest that the Collection unit performs the best in three areas: finance, customer, and innovation and learning. The Revenue Assurance unit performs better in the internal business processes perspective, but its contribution (.23) to the final decision regarding the (overall) best performing unit is fairly similar to the contribution of the Fraud unit (.20). However, decision-makers in the firm clearly consider the contribution of the Collection unit (.57) to be important. Thus, AHP allows managers to identify the Collection unit as, in relative terms, superior, taking into account the unit's achievement in each performance indicator and the relative importance of performance perspectives and indicators. While the weights may seem to imply overemphasis on financial indicators, this state of affairs does not constitute a problem with the method, but rather, a reflection of the idiosyncratic preferences (and consequent performance assessment criteria) of the particular functional area under evaluation. However, the fact that the method forces managers to consider more than one perspective prevents exclusive reliance on a purely financial viewpoint.

Although the consensual process that leads to the final global performance of each functional unit seems to provide a protection against possible individual bias by decision-makers, the literature recommends a check of convergent validity. Smith (1999) argues that researchers can perform such a validity check by eliciting information about overall assessments, rated on a different scale and as a separate question. Thus the Vice-President of Finance rank-ordered the three functional units in terms of the list with the nine performance indicators and their respective values (presented in the original units, neither in the nine-point scale nor as normalized weights). His independent rank-order of the units was the same as that generated by the AHP method – thus providing some evidence of convergent validity. As a footnote, the Vice-President of Finance explained that financial indicators are the easiest and the simplest to collect and that cash inflows are more visible than (reduced) costs or avoided losses – thus favoring the Collection unit over the other two units. He also contended that he ranked the Fraud unit last because its indicators would be more difficult to measure.

5. Conclusions

This study integrates two tools, BSC and AHP, to provide a better assessment of the (relative) performance of three organizational units within a Brazilian telecommunications company. While BSC explicitly incorporates several perspectives (besides the usual financial viewpoint) to organizational performance evaluation, AHP handles multiple perspectives (criteria) and measures (sub-criteria) with distinct degrees of importance, and translates the overall result into a unified metric. AHP circumvents the pitfalls of having managers use a simple, *ad hoc*, weighting approach, and helps them make sense of the multiplicity of performance measures from a balance scorecard. AHP provides more than an ordinal ranking and informs the magnitude of the difference between alternatives, i.e., how a functional area performs better (worse) than another.

As noted above, the study seeks to strengthen the adequacy of the focus group-like approach of (i) eliciting the priority values of perspectives and indicators in a balanced scorecard framework and (ii) weighting performance outcomes. The particular numerical values of weights for perspectives and indicators – and thus final comparative analysis among the units – are specific to the case studied. This specificity of results reflects not only firm characteristics, but also telecommunication industry characteristics, country environment and temporal moment. The values themselves are not a theoretical issue, but rather, a managerial one. However, the procedure to derive relative weights and rankings of alternatives here presented is sufficiently generic to be of utility to other firms, regardless of industry or country.

The application of BSC and AHP methodologies in this particular case – where several rounds of discussion are necessary before managers can reach informed agreement – underscores the fact that organizational performance is a complex and multifaceted phenomenon. Managers must employ a sophisticated framework for the assessment of performance of organizational units – one that explicitly accounts for multidimensionality, incongruence among objectives, and mutually inconsistent evaluations.

The following paragraphs acknowledge some limitations of the study and describe how the authors address them. Firstly, the authors do not perform a sensitivity analysis to verify whether results are reasonably stable — for example, preservation of the same ranking among the current alternatives, with the addition of some other alternatives (functional units) or criteria (performance perspectives and measures). Also, recognizing that a distinct group of managers might provide a distinct evaluation picture, this study employs a multiplicity of managers (16) and hierarchical levels (four) to embrace explicitly different viewpoints and to minimize particularistic assessment. Additionally, the study checks convergent validity by comparing the rankings derived from AHP with the opinion of a senior insider expert (the Vice President of Finance), whose rankings the researchers elicited without his prior knowledge of the final rankings of the other managers. By working with participants from the same department and with similar backgrounds, the study aims to preserve internal validity in terms of "equivalence of groups on participant characteristics and control of extraneous experience or environment variables" (*cf.*, Gliner & Morgan, 2000, p.88).

The risk of self-assessment bias might threaten internal validity and constitutes a limitation of this study: the employees who propose the performance indicators are the same individuals who determine the relative degree of importance of such variables and who assess the relative level of achievement in each indicator. The consequences (good or bad) of the ensuing results also have an effect on managers. Further studies could emphasize internal validity issues by attempting, for instance, to compare a self-assessment performance using the AHP-BSC framework with another measure of performance defined independently by others in the firm or by a third-party. A promising avenue would be Evidential Reasoning (ER), a more sophisticated multi-criteria decision analysis method for self-assessment (Xu & Yang, 2003, Yang & Singh, 1994, Yang & Xu, 2002) that assumes that uncertain subjective judgments satisfy rationality assumptions. The approach borrows from decision science,

artificial intelligence, statistical analysis, fuzzy theory and computer technology to reduce subjectivity in self-assessment analysis (Xu & Yang, 2003).

The dynamic characteristics of the external environment (changes in internal strategy and in external pressures) inherently represent an important obstacle for any performance assessment mechanism. Therefore, any tool for performance evaluation should consider the dynamism of both firm and environment. However, mechanisms based on comparison between prior definition of goals and posterior measurement of indicators usually fail to address the fact that needs and priorities change over time. However, adjustment of priorities over time was outside the scope of this study. Importantly, the joint procedure using AHP-BSC aims to identify the best performing unit in a period, taking into account current prioritization of indicators and perspectives. This fact underscores another (albeit polemic) advantage of the procedure, which is the evaluation of past performance in the context of current prioritization criteria. Thus, the time horizon gap avoids, for instance, a bad performing unit, evaluated under current standards, as being identified as a good performing unit simply because it met goals defined for a previous period. The AHP procedure is not a dynamic mechanism per se, but is able to adapt to the effects of a changing environment, which is the case in the volatile telecommunications industry.

Limitations notwithstanding, the study's contributions are important. Leung et al. (2006, p. 683) suggests that, "although the conceptual framework of the BSC has been widely accepted in the business community, the proper method of implementing the framework remains an issue." Although the present study's contribution to the advancement of the theory is modest, the approach described can aid in the comprehension of AHP and in surmounting the challenges posed by its application as a multi-criteria decision-making tool in real-world BSC performance assessment contexts. Managers reported that the endeavor as a whole, although demanding effort and time, helped them see the big picture and identify synergies and possible areas for global improvements, rather than focusing narrowly on a personal agenda. In fact, the joint consideration of several department-wide goals and the contributions of each functional unit – along several performance perspectives and metrics – leads managers to organize common efforts better and identify corrective actions that look beyond short-term financial results. The focus-group design illuminated managers' understanding of the business and its outcomes, while at the same time fostering teamwork.

The approach of using group discussions to seek consensus for the AHP input constitutes a contribution to better application of theoretical models in real-world situations. Furthermore, although several studies present theoretical arguments and foundations to AHP and BSC, few provide explicit advice when it comes to implementing such concepts in a real organization. This study is an attempt to combine theory with practice. A balanced approach (BSC) to the assessment of performance is important because multiple perspectives (dimensions) and multiple measures (indicators) exist via which the investigator can conceive of and rate the performance of functional areas. Given that managers may not attribute the same degrees of performance to all dimensions and to all indicators – and, besides, a functional area may not perform equally well (poorly) in all indicators (and dimensions) – a decision-making tool (AHP) that accounts for this complexity and helps rank-order the alternatives under evaluation appears useful. As demonstrated, the integrated approach (BSC and AHP) employed here presented convergent validity, providing a fine-grained picture of performance (through the in-depth discussions among managers) and overall assessment. The interactive and iterative process employed in this study has the additional advantage of enabling managers to apprehend the diverse perspectives of performance assessment and to understand possible trade-offs. Managers receive insights as to which functional areas are doing well and which require more attention. Results can serve as input, for example, to shape bonus distribution, incentive schemes, and discussion of possible reasons for performance gaps and successes.

References

Akkermans HA, van Oorschot KE. Relevance assumed: a case study of Balanced Scorecard development using system dynamics. J Oper Res Soc 2005;56(8):931–41.
Ambler T. Marketing metrics. Bus strategy rev 2000;11(2):59–66.
Bana e Costa C, Vansnick J. Macbeth: an interactive path towards the construction of cardinal value functions. Int Trans Oper Res 1994;1(4):489–500.
Banker RD, Chang H, Pizzini MJ. The Balanced Scorecard: judgmental effects of performance measures linked to strategy. The Acc Rev 2004;79(1):1-23.
Barney J. Gaining and sustaining competitive advantage. 4th ed. Upper Saddle River: Prentice Hall; 2010.
Bernhard R, Canada J. Some problems in using benefit/cost ratios with the Analytic Hierarchy Process. The Eng Eco 1990;36(1):56–65.
Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation perspective. Psychol Bull 1991;110(2):305–14.
Brenes E, Haar J, Requena B. Latin America: environmental and firm-level challenges. J Bus Res 2009;62(9):849–53.
Cameron K. Effectiveness as paradox: consensus and conflict in conceptions of organizational effectiveness. Manage Sci 1986;32(5):539–53.
Chakravarthy B. Measuring strategic performance. Strateg Manage J 1986;7:437–58.
Chan Y. An Analytic Hierarchy framework for evaluating Balanced Scorecards of healthcare organizations. Can J Adm Sci 2006;23(2):85-104.
Chan YCL, Lynn BE. Performance evaluation and the Analytic Hierarchy Process. J Manag Acc Res 1991;3:57–87.
Chou Y, Lee C, Chung J. Understanding m-commerce payment systems through the Analytic Hierarchy Process. J Bus Res 2004;57(12):1423–30.
DeBusk GK, Brown RM, Killough LN. Components and relative weights in utilization of dashboards measurement systems like the Balanced Scorecard. Br Acc Rev 2003;35(3):215–31.
Diamantopoulos A. Export performance measurement: reflective versus formative indicators. Int Mark Rev 1999;16(6):444–57.
Diamantopoulos A, Kakkos N. Managerial assessments of export performance: conceptual framework and empirical illustration. J Int Mark 2007;15(3):1-31.
Dyer J. Remarks in the Analytic Hierarchy Process. Manage Sci 1990;36(3):249–58.
Fletcher HD, Smith DB. Management for value: developing a performance measurement system integrating Economic Value Added and the Balanced Scorecard in strategic planning. J Bus Strateg 2004;21(1):1-17.
Forman E, Gass S. The Analytic Hierarchy Process – an exposition. Oper Res 2001;49(4):469–86.
Forman E, Peniwati K. Aggregating individual judgments and priorities with the Analytic Hierarchy Process. Eur J Oper Res 1998;108:165–9.
Gliner JA, Morgan GA. Research design and analysis in applied settings: an integrated approach to design and analysis. Mahwah NJ: Lawrence Erlbaum; 2000.
Goodwin P, Wright G. Decision analysis for management judgment. 3rd ed. New York: Wiley; 2000.
Harker P, Vargas L. The theory of ratio scale estimation: Saaty's Analytic Hierarchy Process. Manage Sci 1987;33(11):1383–403.
Ittner CD, Larcker DF, Meyer MW. Subjectivity and the weighting of performance measures. Acc Rev 2003;78(3):725–58.
Javalgi R, Armacost R, Hosseini J. Using the Analytic Hierarchy Process for bank management: analysis of bank selection decisions. J Bus Res 1989;19(1):33–49.
Jovanovic J, Krivokapic Z. AHP in implementation of Balanced Scorecard. Int J Qual Res 2008;2(1):59–67.
Kaplan R, Norton D. The Balanced Scorecard: translating strategy into action. Boston: Harvard Business Press; 1996.
Kaplan R, Norton D. The Balanced Scorecard: measures that drive performance. Harv Bus Rev 1992:71–9. (Jan-Feb).
Keegan D, Eiler R, Jones C. Are your performance measures obsolete? Manage Acc 1989;70(12):45–50.
Kumar M, Bhagwat R. An integrated BSC-AHP approach for supply chain management evaluation. Measuring Bus Excell 2007;11(3):57–68.
Lane E, Verdini W. A consistency test for AHP decision makers. Decis Sci 1989;20(3):575–90.
Leskinen P. Measurement scales and scale independence in the Analytic Hierarchy Process. J Multicriteria Decis Anal 2000;9(4):163–74.
Leung L, Lam K, Cao D. Implementing the Balanced Scorecard using the Analytic Hierarchy Process and the Analytic Network Process. J Oper Res Soc 2006;57(6):682–91.
Liberatore M, Monahan T, Stout D. A framework for integrating capital budgeting analysis with strategy. Eng Econ 1992;38(1):31–42.
Libby T, Salterio S, Webb A. The Balanced Scorecard: the effects of assurance and process accountability on managerial judgment. Acc Rev 2004;79(4):1075–94.
Lipe M, Salterio SE. The Balanced Scorecard: judgmental effects of common and unique performance measures. Acc Rev 2000;75(3):283–98.
Lu M, Madu C, Kuei C, Winokur D. Integrating QFD, AHP and benchmarking in strategic marketing. J Bus Ind Mark 1994;9(1):41–50.
Madu C, Kuei C, Madu A. Establishing priorities for the information technology industry in Taiwan: a Delphi approach. Long Range Plann 1991;24(5):105–18.
Morgan D, Krueger R. When to use focus groups and why. In: Morgan D, editor. Successful focus groups: advancing the state of the art Thousand Oaks: Sage; 1993. p. 3-19.

Neely A. The performance measurement revolution: why now and what next? Int J Oper Prod Manage 1999;19(2):205–28.
Neely A, Gregory M, Platts K. Performance measurement system design: a literature review and research agenda. Int J Oper Prod Manage 1995;15(4):80-116.
Reisinger H, Cravens KS, Tell N. Prioritizing performance measures within the Balanced Scorecard framework. Manage Int Rev 2003;4(43):429–37.
Roy B. Classement et choix en présence de points de vue multiple (la méthode ELECTRE). RIRO 1968;2:57–75.
Roy B, Vincke Ph. Relational systems of preference with one or more pseudo-criteria: some new concepts and results. Manage Sci 1984;30:1323–35.
Saaty T. Group decision making and the AHP. In: Golden B, Wasil E, Harker P, editors. The Analytic Hierarchy Process: applications and studies. Berlin: Springer; 1989. p. 59–67.
Saaty T. Comment on rethinking value elicitation for personal consequential decisions by Wright G, Goodwin P. J Multi-Criteria Decis Anal 1999;8(1):23–4.
Saaty T. Multicriteria decision making: the Analytic Hierarchy Process. Pittsburgh: RWS Publications; 1990a.
Saaty T. An exposition of the AHP in reply to the paper 'Remarks on the Analytic Hierarchy Process'. Manage Sci 1990b;36(3):259–68.
Saaty T. Axiomatic foundation of the Analytic Hierarchy Process. Manage Sci 1986;32(7):741–855.
Saaty T. The Analytic Hierarchy Process. New York: McGraw-Hill; 1980.
Schoner B, Wedley WC. Ambiguous criteria weights in AHP: consequences and solutions. Decis Sci 1989;20(3):462–75.
Shahin A, Mahbod M. Prioritization of key performance indicators: an integration of Analytical Hierarchy Process and goal setting. Int J Prod Perform Manage 2007;56(3):226–40.
Sharma M, Bhagwat R. An integrated BSC-AHP approach for supply chain management evaluation. Measuring Bus Excell 2007;11(3):57–68.
Smith A. Some problems when adopting Churchill's paradigm for the development of service quality measurement scales. J Bus Res 1999;46:109–20.
Smith J, von Winterfeldt D. Decision analysis in management science. Manage Sci 2004;50:561–74.
Varma S, Wadhwa S, Deshmukh SG. Evaluation petroleum supply chain performance: application of Analytical Hierarchy Process to Balanced Scorecard. Asian Pac J Mark Logistics 2008;20(3):343–56.
Venkatraman N, Ramanujam V. Measurement of business performance in strategy research: a comparison of approaches. Acad Manage Rev 1986;11(4):801–14.
Vincke Ph. Multicriteria decision-aid. New York: John Wiley & Sons; 1992.
Voronin A. A method of multicriteria evaluation and optimization of hierarchical systems. Cybern Syst Anal 2007;43(3):384–90.
Xu DL, Yang JB. Intelligent decision system for self-assessment. J Multi-criteria Decis Anal 2003;12(1):43–60.
Yang JB, Singh MG. An evidential reasoning approach for multi-attribute decision making with uncertainty. IEEE Trans Syst Man Cybern 1994;24(1):1-18.
Yang JB, Xu DL. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. IEEE Trans Syst Man Cybern Part A: Syst Hum 2002;32(3):289–304.
Zammuto R. A comparison of multiple constituency models of organizational effectiveness. Acad Manage Rev 1984;9(4):606–16.
Zou S, Taylor CR, Osland GE. The EXPERF scale: a cross-national generalized export performance measure. J Int Mark 2008;6(3):37–58.