

## Accounting and Critical Theories

### TABLE OF CONTENTS / TABLE DES MATIERES

Note: The Accounting and Critical Theories Division did not send us their proceedings. But they did send us the following paper asking that we include it in the proceedings:

**WEIGHING THE BALANCED SCORECARD: AN EXAMINATION OF RELEVANT RESEARCH ..... 2**

Stephen M. Bryan (student)  
Steven A. Murphy  
Sprott School of Business  
Carleton University

## **WEIGHING THE BALANCED SCORECARD: AN EXAMINATION OF RELEVANT RESEARCH**

The Balanced Scorecard (BSC) has evolved from a performance measurement tool to become a widely accepted strategic management system. This paper examines relevant research as it relates to the most pressing methodological and outcome issues surrounding the BSC and concludes that potential issues in each area can be traced back to its concept generation through innovation action research.

### **Introduction**

As their pace of change accelerates as a result of technological and environmental factors, organizations are faced with challenges in measuring performance. The Balanced Scorecard (BSC) enables managers to translate an organization's strategic goals into both financial and non-financial performance measures (Kaplan & Norton, 2001). Since being initially proposed by Robert Kaplan and David Norton in 1992 (Kaplan & Norton, 1992), the purported role of the BSC has grown from simply measuring performance to one aimed at ensuring strategic alignment across the organization. In this respect, it has become an approach to management, not simply just to measurement. While other performance measurement tools exist<sup>1</sup>, the BSC is the most commonly deployed approach. As of 2002, the BSC was in use in approximately half of the Fortune 1000 firms in the United States (Sale & Sale, 2005).

The implementation of a performance measurement system such as the BSC requires organizational commitment and resources. While it is commonly accepted performance measurement can make an important contribution to organizational management, the exact nature of that contribution and the effectiveness of the available tools are not universally agreed upon.

Performance measurement systems, and the BSC in particular, have been the subject of ongoing examination by the practitioner and academic communities over the past decade (Kennerley & Neeley, 2002). The majority of this effort has examined BSC implementation, with results often being found in practitioner publications within the management accounting domain. However, considerable academic research has also been undertaken, although much of its focus has also been on implementation rather than theoretical issues. Studies have sought to more rigorously examine the fundamental model, and, in so doing, have identified perceived flaws. In view of the wide adoption of the BSC construct, these flaws should be the source of practitioner concern and researchers interest; however this does not necessarily appear to be the case.

This paper aims to provide the reader with an understanding of the relevant research on the most pressing methodological and outcome issues surrounding the BSC. In so doing, it does not purport to provide a detailed critique of the BSC, but rather to examine the extent to which researchers have

undertaken such work and, where applicable to analyse the conclusions they have reached. This review is framed within the evolution of the BSC from measurement to management construct and it will examine approaches pursued on three different levels:

- From an implementation perspective, it will identify research aimed at addressing two elements of potential construct bias:
  - Common measures bias.
  - Subjectivity bias in measure weighting and techniques advanced to mitigate these concerns.
- From a construct validity perspective, it will review research into the cause and effect hypotheses underlying the BSC approach.
- From a predictive validity viewpoint, it will examine research into whether or not the BSC achieves its forecast benefits in improving organizational performance.

The remainder of this paper is organized as follows. The next section will outline the fundamentals of the BSC and briefly describe its evolution while the subsequent two sections will examine how researchers have addressed the questions of potential common measures and subjectivity bias. Research investigating the cause and effect hypotheses will then be considered. Prior to concluding, the paper will look briefly at theoretical and empirical efforts aimed at assessing BSC performance and then identify potential research implications and areas for future efforts.

### **Evolution of the Balanced Scorecard**

While financial measures of an organization's performance have long been the backbone of performance reporting, they are focused on revealing what has occurred. By ignoring the value of intangible assets such as goodwill, research and development and human capital, these accounting measures do not necessarily provide sound indicators of future performance (Norreklit, 1999).

The BSC seeks to address the shortcomings associated with using only financial measures by supplementing financial performance indicators with performance criteria from three additional perspectives: the customer, internal business processes and learning and growth (Kaplan & Norton, 1992). The reasons for adopting this broader perspective are as follows:

- Financial measures provide the ultimate result of an organization's performance but are a lag indicator.
- Customer satisfaction measures offer a lagging indication of performance (market share, retention and growth rates, etc) and assist in the forward looking development of the customer value proposition.
- The performance of internal business processes are important leading indicators of future financial and customer performance.
- The learning and growth perspective captures the intangible assets within the organization and provides a leading indicator of future success in internal business processes, customer and financial outcomes (Kaplan & Norton, 2004).

Kaplan and Norton advance the scorecard as providing a balanced view across the perspectives and furnishing decision-makers with the necessary information on both current operating performance and the drivers of future performance (Kaplan & Norton, 1996). When they published their initial BSC paper in 1992, it was this combination of financial and non-financial indicators that distinguished the BSC from other performance measurement approaches (Neeley, 2005).

Kaplan and Norton developed the BSC through case studies with twelve firms (Kaplan & Norton, 1992). Kaplan has subsequently labelled the approach used as being innovation action research (Kaplan, 1998; Norreklit, 2003). This approach involves the documentation of a current shortfall in current practice, the identification of a concept to address the shortfall, and then the ongoing improvement of the concept through publication, teaching and engagement with organizations who then apply the described techniques (Kaplan, 1998).

The 1992 Kaplan and Norton paper addresses the first two elements of the innovation action research paradigm and it initiated both a plethora of writings in academic, practitioner and general publications and the commercialization of the BSC concept. While some academics might question the validity of the marketing effort to inculcate BSC fundamentals into organizations, the approach taken is fully consistent with the innovation action research paradigm proposed by Kaplan (Norreklit, 2003). This said, and as discussed in subsequent sections addressing common measures and subjectivity bias, the practitioner-driven focus of the implementation effort has led to potential shortfalls in methodological soundness, at the very least from an academic perspective.

In consonance with that innovation action research model, the BSC has evolved significantly since the first 1992 Norton and Kaplan article. While the four perspective approach and the basic tool have remain largely unchanged, the two principal proponents have enlarged the role of the BSC from one which focused almost exclusively on performance measurement to one whose primary purpose is to ensure alignment of the company's ongoing actions with its overall strategy. With this enlargement of the initial concept, came four supporting processes: translating the vision, communicating and linking, business planning, and feedback and learning (Kaplan & Norton, 1996). As with the 1992 article, the application of the BSC as a strategic management system was borne from Kaplan and Norton's observations of how organizations were implementing management techniques to fill an identified need.

The final step to date in the BSC's evolution was the formulation of the concept of a strategy map which more explicitly enunciates the cause and effect relationships inherent in the BSC (Kaplan & Norton, 2001). In this regard, the cause and effect hypotheses seems to have moved, at least in Kaplan and Norton's writings, from being one element of the BSC construct to being its primary foundation. It appears this evolution has, until recently, largely been accepted by researchers whose focus has been on the performance measurement aspects of the BSC. This leaves open, at least to some degree, a question as to the BSC's construct validity. The extent to which researchers have examined the validity of the assumptions underlying the cause and effect models will be examined later in this paper.

Over the past decade, the BSC has evolved from a measurement system that seeks to provide both leading and lag indicators of an organization's performance, to a management system serving as a fundamental organizing framework (Kaplan & Norton, 2001). The inclusion of the strategy maps element further extended the initial model to provide it a role in managing organizational transformation (Kaplan & Norton, 2001a).

With a limited number of exceptions (e.g., Norreklit, 2000), the theoretical underpinnings of this transition from measurement system to management system appears to have largely escaped critical assessment by the academic community. In an effort to address issues such as "does the BSC work?" and "how can it be made to work better?" questions related to the theoretical soundness of the construct appear to remain unanswered and, to a certain extent, unasked.

### **Common Measures Bias**

To fully exploit the BSC's potential, Kaplan and Norton encourage organizations to develop scorecards at different levels of the organization. In so doing, they note such scorecards are likely to be a

combination of both generic measures that are common across the organization, and those that are specific to the division or operating unit (Kaplan & Norton, 1996). This creates circumstances wherein a senior manager might be reviewing performance scorecards of subordinate divisions that are characterised by both common and specific measures.

In referring to work by Slovic and MacPhillamy (1974), Lipe and Salterio (2000) note previous studies have shown organizations and individuals place greater weight on common information when making comparisons between two individuals, groups, etc. However, significant dominance of common measures could undermine the utility of the differentiated measures of the BSC, and suggest that the use of only common measures across divisions would achieve largely the same result as the more complex and effort-intensive BSC.

Lipe and Salterio (2000) examined the potential impact of common measures bias in the BSC by conducting an experiment with 58 MBA students. The students were provided a scenario outlining the operational strategies and BSC performance indicators of two divisions of a hypothetical clothing manufacturing firm. The scorecards contained 16 separate measure BSCs for each division (four per perspective, two of which were common, two of which were specific). Performance data was provided, and the participants were asked to rate the performance of the two divisional managers. The goal of the experiment was to determine whether performance on common and unique measures effects the evaluation of the division managers (Lipe & Salterio, 2000).

Lipe and Salterio (2000) used a 2 X 2 X 2 repeated measures ANOVA to test their hypothesis that both common and unique measures contributed to the evaluation. The experiment's results do not support the hypothesis, but rather portray the dominance of common measures over unique measures.<sup>2</sup> This suggests managers, by employing simplifying cognitive strategies, largely ignore one of the key elements of the BSC concept, namely performance measures unique to the strategic objectives of the business unit (Lipe & Salterio, 2000). This simplification technique undermines the scorecard's effectiveness in distinguishing performance between measures whose importance may vary by division. Given that it may well be the unique measures that are the most important indicators, this sets the conditions for potentially improperly interpreted results and inaccurate conclusions.

Roberts, Albright and Hibberts (2004) hypothesized that presenting the BSC in a disaggregated format would result in evaluations reflecting both common and unique measures, thereby addressing the common measures bias identified by Lipe and Salterio (2000). They tested their hypothesis by repeating the Lipe and Salterio experiment, but rather than comparing the two scorecards, the participants were required to make an individual assessment of each manager by assigning an evaluation against weighted criteria. The weighted criteria scores were then to produce a final result.

Roberts, Albright and Hibberts (2004) labelled this information processing strategy of evaluating performance measures separately and then aggregating the separate results using pre-assigned weights as a 'disaggregated/mechanically aggregated' scorecard. It produced results that support the hypothesis that both common and unique measures contribute to the evaluation<sup>3</sup> (Roberts et al., 2004). This would appear to be an effective means of addressing the common measures bias within the BSC construct.

The tendency to conduct relative rather than absolute comparisons could result from effort conservation techniques. When combined with potential concerns as to the quality of non-traditional measures, this could cause managers to ignore unique BSC measures (Libby, Salterio & Web, 2004). Libby et al. (2004) repeated elements of the Lipe and Salterio (2000) experiment and found that providing process accountability by requiring managerial justification of evaluations or quality assurance through third party reports increased the use of unique measures (Libby et al., 2004).

Notwithstanding the potential mitigating strategies outlined above, common measures bias has important implications for the validity of the BSC. Firstly, if unique measures are disregarded or underweighted in the *ex post* performance evaluation of business units, this is likely to influence the *ex ante* decision making strategy of that unit (Lipe & Salterio, 2000). Secondly, the common measures are mostly likely to be traditional financial measures that are lagging indicators of performance; the unique measures are more likely to be non-traditional, leading indicators (Roberts et al., 2004). Disregarding these unique measures likely returns an organization to a pre-BSC state of relying almost exclusively on financial, lag indicators but possibly with a false sense of having a more nuanced view of its performance.

While Kaplan and Norton do not acknowledge the potential challenges raised by common measures bias, the three studies identified in this paper confirm its potentially negative effects on the purported benefits of the BSC.

The effectiveness of these mitigating strategies is supported by empirical evidence but each requires effort on the part of the organization. In the case of the Roberts et al. (2004) suggestion to over-weight the unique measures to offset the reliance on common ones, additional questions vis-a-vis subjectivity are created.

### **Measure Selection and Subjectivity Bias**

Given concerns over measure reliability, managers often experience difficulties in incorporating BSC subjective measures into the decision-making process (Sale & Sale, 2005). While Kaplan and Norton have emphasized the importance of choosing measures that reflect the specifics of the organization (Kaplan & Norton, 1996), the challenge of measure selection is one of the principal obstacles to effective BSC implementation (Clinton, Webber, & Hassell, 2002). Clinton et al. (2002) cite Frigo and Krumwiede (1999) in noting the inadequacy of performance metrics in many scorecards.

Much of the BSC related literature (and that describing other performance measurement approaches) has remained at the framework level, resulting in only broad guidance with respect to the actual development of suitable measures (Neeley, Mills, Platts, Richards, Gregory, Bourne & Kennerley, 2000). One gains little useful insight into measure development from the writing of Kaplan and Norton and others. This may well be a result of the innovation action research methodology applied during the conceptual development of the BSC. A good deal of BSC literature, particularly in the practitioner domain, is case study based, and focuses on describing the imperatives leading to BSC adoption and the approach taken during implementation. It is not surprising that, with an emphasis on widely applicable lessons, such writings would pay only limited attention to the case specific mechanics of measure development.

Ittner, Larcker and Meyer (1999) summarize existing research by noting that if one assumes reliability is an important factor in choosing performance measures, greater weight will be placed both on quantitative rather than qualitative measures and on measures that are aggregations of multiple indicators rather than single indicator based measures. However, as noted earlier in this paper, such a situation would lead to an overweighting of lagging financial indicators and an underweighting of the leading but more subjective indicators associated with the other three BSC perspectives.

Ittner, Larcker and Randall (2003) examined measure subjectivity in the conduct of a study using case data from senior management evaluations and incentive plans of a large North American retail financial firm. Amongst other techniques, they regressed the overall par score developed from the organization's scorecard against the final managers' evaluations and incentive plan rewards. They noted that subjectivity in weighting measures in the scorecard enabled supervisors to ignore many of the

performance measures and to emphasize financial measures over those of the other perspectives. They also observed that the subjectivity present enabled those supervisors to import other performance factors not identified in the BSC into the overall evaluation (Ittner et al., 2003).

Information overload is another factor impacting both organizational performance and the extent to which the BSC provides an effective management tool (McWhorter, 2003). In a survey of approximately 1,500 management accountants, McWhorter (2003) observes that BSC users report the information they receive is more strategically focused and supportive of decision-making than non-BSC users, however, they also note that users report having a slightly higher number of measures.

To address both measure subjectivity and the risk of information overload, a number of authors have recently proposed the use of Analytical Hierarchy Process (AHP) as means to assist in measure selection and weighting (see Searcy, 2004; Clinton, et al., 2002, Sale & Sale, 2005; Chan, 2006). AHP uses a series of pairwise comparisons in which users assess the relative dominance of two items where dominance can be expressed in any relevant criterion (Saaty, 1994). The resultant eigen values are used to establish the priorities of the factors involved. In this manner, AHP can be used to minimize the subjectivity in both choosing and in weighting measures (Clinton et al., 2002). Internal rating consistency can be evaluated to ensure result validity (Chan, 2006).

Using alignment with the corporate strategy as a criterion, Searcy (2004) used AHP to construct BSC performance factors for six firms. Whereas Searcy used AHP in BSC construction, Chan (2006) used it to examine the relative performance of eight community care hospitals in Ontario against existing performance criteria. This was achieved by determining, through a series of pairwise comparisons, the priority or weight of each performance perspective and associated indicators. These weights became the factors against which results were multiplied to produce a weighted performance result.

Both the Searcy (2004) and Chan (2006) studies demonstrate the potential utility of AHP as a means of addressing possible measure subjectivity bias. The Chan (2006) study takes an important step by identifying that individual evaluators can assess performance by assigning different weights to criteria in line with their particular interests but within a mechanism that reduces inherent subjectivities. In the case of the example cited by Chan (2006), this would enable the various stakeholders (the public, hospital administrators, regional health authorities, etc) to assess hospital performance based on commonly agreed to measures but within individually determined weighting. The aggregation of these weighted assessments would provide a more comprehensive stakeholder view than might be achievable through other, less rigorous means.

While AHP represents an improvement over the ad-hoc weightings and subjective criteria selection processes that can plague multi-attribute decision models, it is time-consuming and potentially complicated. In addition, as noted by Chan (2006), the behavioural impact of AHP decisions on managers has yet to be studied. Nevertheless, AHP appears to be a promising approach to reducing the subjectivity bias inherent in the selection and weighting of BSC measures.

### **Cause and Effect**

The previous two sections of this paper have examined some of the relevant research as it applies to addressing two important implementation issues associated with the BSC, namely common measures and subjectivity bias concerns. This section will address research into a more fundamental aspect of the BSC paradigm, namely that of the extent to which causality exists between the various elements of the scorecard.

Commencing with their 1996 book, Kaplan and Norton (1996) emphasize the causal link between the four performance perspectives in the BSC. The measures of organizational learning and growth become a leading indicator of internal business processes which in turn drive the measures of the customer perspective and so on. This causal link represents the evolution of the BSC from its initiation as a performance measurement tool to one of being a strategic management system (Kaplan & Norton, 1996a). It is also essential to the strategic view of the BSC that non-financial measures can be used to predict future financial results (Norreklit, 2000).

Norreklit (2000) critically examined the BSC model from an assumptions perspective and concluded that, notwithstanding Kaplan and Norton's assumption of a cause and effect chain, such an assumption is not theoretically supported. He notes the lack of a time dimension within the scorecard (other than broad classifications of leading and lagging indicators) and suggests that there is a logical rather than causal relationship between the performance outcomes. He cites, as an example, that high levels of customer satisfaction do not necessarily lead to positive financial results (Norreklit, 2000). In a subsequent article, he observes that Kaplan and Norton have erred by interpreting factor covariance as evidence of cause and effect (Norreklit, 2003).

While other authors acknowledge both the central role of the cause and effect assumption in the post 1996 versions of Kaplan and Norton's writings and Norreklit's criticism as to its lack of validity, (see Braam & Nijssen, 2004, Davis & Albright, 2004, Malina & Selto, 2001), few seem to have pursued these concerns.

Othman's research (2006) examines the extent to which Malaysian companies build their scorecards on a strategy-founded causal model. While the study is hindered by a relatively small sample (36 respondents), it provides evidence that causal assumptions positively impact BSC implementation. The absence of a causal model resulted in difficulties developing non-financial (leading) indicators. Othman's (2006) findings are consistent with those of a study of U.S. retail banking (Davis & Albright, 2004), but such findings do not directly address Norreklit's (2003) concerns as to whether or not the BSC is, as Kaplan and Norton propose, a cause and effect model.

### **Balanced Scorecard Predictive Validity**

As previously noted, much of the BSC related literature addresses practitioner level issues related to implementation. Few studies appear to address the more fundamental question as to whether the BSC is effective in improving organizational performance. Given the breadth of BSC adoption and the resources required to implement it, the limited research on its predictive validity is somewhat surprising. In addition, the findings are not consistent across studies although, in consideration of the limited sample sizes, specific questions and varying research approaches described in the following paragraphs, this result is less surprising.

Hoque and James (2000) used survey data from Australian manufacturing firms to assess the relationship between organizational size, product life-cycle stage and market position and BSC usage and between BSC usage and organizational performance. The study found a positive relationship between BSC usage and organizational performance, but the authors acknowledge their measure of BSC usage does not assess the extent to which the BSCs in question were used as a management tool or merely as a measurement mechanism (Hogue & James, 2000).

Malina and Selto (2001) focused on BSC management control benefits of a scorecard in a large US-based company. Data was gathered through qualitative interviews with a large number of involved

managers. They found a positive impact in terms of the development, communication and implementation of strategy and evidence that managers re-aligned their effort and resources to reflect BSC measures.

Somewhat contradictory results are provided by Ittner, Larcker and Randall (2003) who, using data from US financial services firms, found that the implementation of performance measurements systems do result in greater measurement satisfaction levels. Tellingly, they find no relationship between such systems and economic performance (Ittner, Larcker & Randall, 2003). This calls into question the linkage between measure satisfaction and financial performance espoused by, amongst others, Kaplan and Norton (Kaplan & Norton, 2001).

Davis and Albright (2004) followed a quasi-experimental approach to investigate whether four US bank branches using a BSC outperformed five similar branches that did not. The study's statistically significant findings support the hypothesis of superior financial performance by the test group when compared to a control group. Unfortunately, data access issues precluded this study from examining the underlying non-financial measure performance, and hence this study does not address the cause and effect issue identified in the previous section of this paper.

In a survey-based examination of 100 Dutch companies and in consonance with the findings of Malina and Selto (2001), Braam and Nijssen (2004) found a positive relationship between BSC use and organizational performance when the scorecard is directly linked with the organization's strategy. However, they observed a negative result when the BSC was being used solely as a measurement tool.

This overview of existing research suggests mixed results in the empirical evidence examining the BSC's contribution to better organization performance. While there is evidence that the BSC leads to more effective performance measurement, it is not clear that this has been consistently translated into improved financial performance. As noted by Braam and Nijssen (2004), the key ingredient appears to be the extent to which the scorecard is linked to the organization's strategy. However without more focused research into this linkage, the researcher and the practitioner are faced with inconclusive results in terms of the BSC's predictive validity.

One possible explanation for such results would be that the predictive validity of the BSC is not as great as its proponents would suggest, and accordingly, its benefits to organizations may be less than expected. However, before accepting such a premise, one should critically consider the nature and extent of the BSC effectiveness-related research completed to date. In view of the limitations of much of this research, it would appear to be premature and unwise to reach conclusions. This does however establish the requirement for a more comprehensive research agenda.

### **Research Implications**

This review of relevant BSC-related research identifies a number of implications for future research, five of which are highlighted below. Inherent in each is the need for more commonly accepted definitions as to what constitutes an implemented BSC (Ittner et al., 2003). The lack of definition specificity undoubtedly contributes to the range of results generated. The applicability of each of the studies noted is limited by situational specifics, sample size or the study's focused nature, making generalization of the results problematic at best. It is therefore an opportune time to examine larger and more robust samples, to conduct more nuanced and probative experiments and to move beyond unique circumstances to those of a more widely applicable nature.

A key implication from the current state of research is the relative dearth of findings related to the effectiveness of the BSC in improving performance. However almost 15 years after the first BSC article,

and approximately a decade since Kaplan and Norton advocated it as a strategic management system, it is time to systematically examine its effectiveness as a control model. More specifically, research should focus on the extent to which, in line with its cause and effect assumptions, the BSC can be used to predict future performance (Neeley et al., 2005). Given the wide adoption of the BSC across a range of for-profit, not-for-profit and public sector organizations, the potential benefits of a more rigorous examination of the BSC's predictive validity extend well beyond the academic community.

It is also conceivable that the causal model restricts the applicability of the BSC to static and casual circumstances, found in traditional organizational structures and management settings. The BSC may be less useful in more dynamic environments characterised by recursive rather than causal relationships (Neely, 2005). Acceptance of this hypothesis would suggest that the BSC fits certain operating circumstances but not others, and would have limitations in terms of both its theoretical and practical application. Such thinking, however, is at odds with the positioning of the BSC, by both its commercial and academic proponents, as a universally applicable management and control system. Research addressing the universality of the BSC construct would address this issue and may provide considerable insight into the mixed results obtained to date on the construct's predictive validity.

The extent to which the BSC can be employed to advance organizational change is a largely unexplored aspect of its use (Kasurinen, 2002), but one which should be of considerable interest to both the academic and practitioner communities. In view of the extent to which the BSC has been marketed as a management system, it is surprising that its potential application as a change management tool has not been the subject of greater scrutiny. Assuming the model's fundamental cause and effect relationships are valid, one could hypothesise that one of the BSC's greatest strengths would lie in its potential to assist organizations in understanding underlying change factors and subsequently managing their effects during periods of organizational change and renewal.

Finally, little appears to have been published on the influence of BSC measures and results on both *ex ante* behaviours and *ex post* behaviours of managers. The influence of the BSC paradigm on individual and collective behaviours and on organizational culture should however be linked with its utility as a management control system. Further exploration of these aspects would greatly enhance understanding of the role a BSC plays in the strategic management of the organization and assist both researchers and managers in measuring the direct and indirect impacts of BSC implementation on adopting organizations.

## Conclusions

The BSC was developed in response to the need to improve the performance measurement function within organizations; it has subsequently evolved into a strategic management system that seeks to align actions with strategic objectives. The requirement for success in non-financial measures as a pre-condition to attaining the financial objectives is a primary BSC tenet. These cause and effect relationships remain a fundamental assumption of the BSC paradigm, notwithstanding an apparent lack of evidence supporting their validity.

A key consideration in understanding the development of BSC thinking is to acknowledge its birth through innovation action research. The BSC is a product of case study observation of capability gaps, the development and then evolution of techniques to address those gaps, and finally the wide dissemination of those techniques through many means – academic research being but one of those means. Inherent in its foundations in innovation action research is the fact that the BSC is not theory-based. Indeed, its primary theoretical underpinning – the cause and effect relationships between the four perspectives – appear to have developed as the model evolved from being a performance measurement tool to a strategic management approach. While not inherently flawed, this development process should

give both academic researchers and executives of implementing organizations cause to pause and consider the construct's underpinnings.

This paper has examined relevant research in the implementation-related areas of common measure and subjectivity bias. This research is largely focused on identifying appropriate solutions to these potential implementation problems, albeit solutions that are likely to further add to the complexity of, and effort required for, implementation. The need for this bias mitigation research highlights the assessment that that much of the BSC literature, especially that of its two primary authors (Kaplan & Norton, 1992), remains at the framework level. While Kaplan and Norton intersperse their prescriptions with case study descriptions, they offer little in the way of implementation details to resolve bias related issues. In this respect, the gap between the academic and practitioner literature is dangerously wide.

While research on implementation issues is plentiful, few researchers have responded to the criticisms of Norreklit (2000) that the causality assumption underpinning the BSC lacks theoretical validity. Indeed only recently have researchers begun to examine the extent to which the presence of a causality model influences BSC effectiveness. This represents a significant research shortfall and one that poses serious questions for the validity of the BSC construct as a strategic management system. In this regard, the majority of the management accounting research community appears to have accepted the BSC's evolution from performance measurement to strategic management system without questioning the validity of that transformation.

Finally, the paper has provided an overview of recent research into the predictive validity of the BSC. In this domain, the findings have been somewhat contradictory and much of the effort to date has lacked comprehensiveness. There are, however, indications the research community is beginning to critically examine the extent to which the purported benefits of the BSC are realizable within organizations. While the initial indications are positive, much more remains to be done in this area. Implementation of the BSC requires considerable organizational effort and it is perhaps regrettable that a corresponding commitment has yet to be made to validate that such effort is reaping the promised rewards.

## References

- Braam, G.J.M., & Nijssen, E.J., Performance Effects of using the Balanced Scorecard: A Note on the Dutch Experience”, *Long Range Planning*, 37, 2004, 335-349.
- Chan, Y.C.L., “An Analytic Hierarchy Framework for Evaluating Balanced Scorecards of Healthcare Organizations”, *Canadian Journal of Administrative Sciences*, 23, 2006, 85-104.
- Clinton, B.D., Webber, S.A., & Hassell, J.M., “Implementing the Balanced Scorecard Using the Analytic Hierarchy Process”, *Management Accounting Quarterly*, 3, 2002, 1-11.
- Davis, S., & Albright, T., “An Investigation of the Effect of Balanced Scorecard Implementation on Financial Performance”, *Management Accounting Research*, 2004, 15, 135-153.
- Frigo, M.L., & Krumwiede, K.R., “Balanced Scorecards: A Rising Trend in Strategic Performance Measurement”, *Journal of Strategic Performance Measurement*, February-March, 1999, 44, 42-48
- Hogue, Z., James, W., “Linking Balanced Scorecard Measures to Size and Market Factors: Impact on Organizational Performance”, *Journal of Management Accounting Research*, 2000, 12, 1-17.
- Ittner, C.D., Larcker, D.F., & Meyer, M.W., “Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard”, *The Accounting Review*, 78, 2003, 725-758.
- Ittner, C.D., Larcker, D.F., & Randall, T., “Performance Implications of Strategic Performance Measurement in Financial Services Firms”, *Accounting, Organizations and Society*, 2003, 28, 715-741.
- Kaplan, R.S., “Innovation Action Research”, *Journal of Management Accounting Research*, 1998, 10, 89-119.
- Kaplan, R.S., & Norton, D.P., “The Balanced Scorecard: Measures that Drive Performance”, *Harvard Business Review*, January-February, 1992, 71-79.
- Kaplan, R.S., & Norton, D.P., “Putting the Balanced Scorecard to Work”, *Harvard Business Review*, September-October, 1993, 134-147.
- Kaplan, R.S., & Norton, D.P., 1996 *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Publishing.
- Kaplan, R.S., & Norton, D.P., “Linking the Balanced Scorecard to Strategy”, *California Management Review*, 39, 1996a, 53-79.
- Kaplan, R.S., & Norton, D.P., “Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part 1” *Accounting Horizons*, 15, 2001, 87-104.
- Kaplan, R.S., & Norton, D.P., “Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part 2”, *Accounting Horizons*, 15, 2001a, 147-160.
- Kaplan, R.S., & Norton, D.P., “The Strategy Map: Guide to Aligning Intangible Assets”, *Strategy & Leadership*, 32, 2004, 10-17.

- Kaplan, R.S., & Norton, D.P., "Measuring the Strategic Readiness of Intangible Assets", *Harvard Business Review*, February, 2004, 52-64.
- Kasurinen, T., "Exploring Management Accounting Change: The Case of Balanced Scorecard Implementation", *Management Accounting Research*, 13, 2002, 323-343.
- Kennerley, M., & Neeley, A., "A Framework of Factors Affecting the Evolution of Performance Measurement Systems", *International Journal of Operations and Production Management*, 22, 2002, 1222-1245.
- Libby, T., Salterio, S.E., & Webb, A. "The Balanced Scorecard: The Effects of Assurance and Process Accountability on Managerial Judgement", *The Accounting Review*, 79, 2004, 1075-1094.
- Lipe, M.G., & Salterio, S.E., "The Balanced Scorecard: Judgemental Effects of Common and Unique Performance Measures", *The Accounting Review*, 75, 2000, 283-298.
- Malina, M.A., & Selto, F.H. "Communicating and Controlling Strategy: An Empirical Study of the Effectiveness of the Balanced Scorecard", *Journal of Management Accounting Research*, 13, 2001, 47-90.
- McWhorter, L.B., "Does the Balanced Scorecard Reduce Information Overload?", *Management Accounting Quarterly*, 4, 2003, 23-27.
- Neely, A., "The Evolution of Performance Measurement Research", *International Journal of Operations and Production Management*, 25, 2005, 1264-1277.
- Neeley, A., Gregory, M., & Platts, K., "Performance Measurement System Design: A Literature Review and Research Agenda", *International Journal of Operations and Production Management*, 25, 2005, 1228-1263.
- Neeley, A., Mills, J., Platt, K., Richards, H., Gregory, M., Bourne, M., & Kennerley, M., "Performance Measurement System Design: Developing and Testing a Process-Based Approach", *International Journal of Operations and Production Management*, 20, 2000, 1119-1145.
- Norreklit, H., "The Balance on the Balanced Scorecard: A Critical Analysis of Some of its Assumptions", *Management Accounting Research*, 11, 2000, 65-88.
- Norreklit, H., "The Balanced Scorecard: What is the Score?. A Rhetorical Analysis of the Balanced Scorecard", *Accounting, Organizations and Society*, 28, 2003, 591-619.
- Othman, R., "Balanced Scorecard and Causal Model Development: Preliminary Findings", *Management Decision*, 44, 2006, 690-702.
- Roberts, M.L., Albright, T.L., & Hibbets, A.R. "Debiasing Balanced Scorecard Evaluations", *Behavioural Research in Accounting*, 16, 2004, 75-88.
- Saaty, T.L., "How to Make a Decision: The Analytic Hierarchy Process", *Interfaces*, 24, 1994, 19-43.
- Sale, R.S., & Sale, M.L., "Lending Validity and Consistency to Performance Measurement", *Managerial Auditing Journal*, 20, 2005, 915-927.

Searcy, D.L., “Aligning the Balanced Scorecard and a Firm’s Strategy Using Analytic Hierarchy Process”, *Management Accounting Quarterly*, 5, 2004, 1-10.

Slovic, P., & MacPhillamy, D. “Dimensional Commensurability and Cue Utilization in Comparative Judgement”, *Organizational Behaviour and Human Performance*, 11 172-194.

---

<sup>1</sup> See Neeley, Gregory & Platts (2005) for an overview of other performance measurement approaches

<sup>2</sup> The regression analysis of the differences in managerial evaluations against common and unique measures produced a common measures co-efficient of 10.87 (t=3.28, p<0.01) and a unique measures co-efficient of 0.08 (t=0.02, p>0.10) (Lipe and Salterio, 2000).

<sup>3</sup> The regression analysis of the differences in managerial evaluations against common and unique measures produced a common measures co-efficient of 5.18 (t=3.63, p<0.001) and a unique measures co-efficient of 8.00 (t=5.67, p<0.001) (Roberts et al., 2004).