

Performance Pay and Managerial Experience in Multitask Teams: Evidence from within a Firm

Rachel Griffith, *Institute for Fiscal Studies
and University College London*

Andrew Neely, *University of Cambridge and
the Advanced Institute of Management Research*

This article exploits a quasi-experimental setting to estimate the impact that a commonly used performance-related pay scheme had on branch performance in a large distribution firm. The scheme, which is based on the Balanced Scorecard, was implemented in all branches in one division but not in another. Branches from the second division are used as a control group. Our results suggest that the Balanced Scorecard had some impact but that it varied with branch characteristics, and, in particular, branches with more experienced managers were better able to respond to the new incentives.

I. Introduction

Many organizations provide some form of performance-related pay to managers and workers, and its use has increased over time (see, e.g., Lazear and Shaw 2007; and Lemieux, Macleod, and Parent 2007). What form

We would like to thank the anonymous firm for use of the data and Laura Abramovsky, Rupert Harrison, Gareth Macartney, Imran Rasul, Kathryn Shaw, Helen Simpson, and seminar participants at the National Bureau of Economic Research, Cass Business School, London Business School, and University College London for helpful comments. The analysis contained in this article was funded by the Economic and Social Research Council (ESRC) and Engineering and Phys-

[*Journal of Labor Economics*, 2009, vol. 27, no. 1]
© 2009 by The University of Chicago. All rights reserved.
0734-306X/2009/2701-0003\$10.00

should such incentives take? This is the topic of many papers in the economic and management literatures. A well-known theoretical result (Holmstrom 1982; Baker, Jensen, and Murphy 1988; Holmstrom and Milgrom 1991; Baker 1992) suggests that workers should not be made jointly responsible for single tasks because sharing responsibility increases the total risk that each worker faces of successfully completing the task without increasing the benefit. Holmstrom and Milgrom also suggest that tasks should be grouped together based on the cost of measuring and rewarding performance. Some workers should do the easy-to-measure tasks, and their pay should be contingent on performance, while other workers should focus on hard-to-measure tasks and receive fixed wages. This is because, if a worker has both easy and hard-to-measure tasks, they will concentrate on the easy-to-measure tasks at the expense of the hard-to-measure tasks.

However, one of the most widely used measurement and incentive schemes—the Balanced Scorecard—runs counter to these results. The Balanced Scorecard was introduced by Kaplan and Norton (1992), and variants of it have been adopted by many large firms and organizations across the globe.¹ It was one of the first of a new type of performance-related pay scheme, now more common, that makes use of the greater amount of information firms have due to advances in technology (see, *inter alia*, Lazear 2001; Bresnahan, Brynjolfsson, and Hitt 2002; Acemoglu et al. 2007; and Lazear and Shaw 2007). The idea behind the Balanced Scorecard is that managers can improve performance by monitoring and rewarding a range of linked activities (usually 15–20 covering four different perspectives—financial, customer, internal, and innovation/learning). These activities are chosen to reflect both current actions that affect future performance and the outcomes of past actions. Performance is typically evaluated at the group level (e.g., business unit, branch, or team) and incentive payments are often based on group performance. As well as forming the

ical Sciences Research Council (EPSRC) through Advanced Institute of Management Research (AIM). All errors and omissions remain the responsibility of the authors. Contact the corresponding author, Rachel Griffith, at rgriffith@ifs.org.uk.

¹The most recent evidence is contained in Neely, Yaghi, and Youell (2008). Other data are available in the Balanced Scorecard Hall of Fame at http://www.bscol.com/pdf/BSCHoF-Membersby_Industry_2000-2005.pdf; or see Gates (1999), who surveys 113 “leading” U.S., European, and Asian companies and finds that 81% of respondents use a strategic performance measurement system. Maisel (2001) samples 1990 U.S. management accountants and finds that 47% of respondents use a strategic performance measurement system. Rigby (2001) surveys 214 North American firms and finds that 44% of organizations use the Balanced Scorecard; Speckbacher, Bischof, and Pfeiffer (2003) estimate that 26% of firms in Germany, Switzerland, and Austria use the Balanced Scorecard; Marr et al. (2004) find that 35% of North American organizations use it.

basis for a performance-related pay scheme, the Balanced Scorecard also provides information to managers and workers: “The Balanced Scorecard is like the dials in an airplane cockpit: it gives managers complex information at a glance” (Kaplan and Norton 1992, 71).

Besides running counter to results from the economics literature, the use and effectiveness of the Balanced Scorecard has been criticized in the management literature.² In particular, Jensen (2001) argues that the Balanced Scorecard will do nothing to solve the agency problem. It provides no information on how managers and workers should trade off different objectives, so will fail as an incentive mechanism, although he accepts that it may provide useful information to managers, for example, on the company’s strategy and drivers of value.

With all these arguments against the Balanced Scorecard, why have so many firms adopted it? Has it been an effective incentive scheme? Does it provide useful information to managers? The rapid uptake of the Balanced Scorecard has been largely fueled by high-profile success stories in other firms, yet there is little rigorous empirical evidence on how well the scheme works.³

In this article we exploit a quasi-experimental setting within a single firm in order to investigate the effectiveness of the Balanced Scorecard. The firm is a large multinational that is listed on the London Stock Exchange. It implemented the Balanced Scorecard in one division with the explicit objective to see how well it would work before rolling it out across the firm. This allows us to overcome one of the main problems in evaluating incentive schemes—the fact that who adopts the incentive mechanisms is endogenous—by comparing monthly performance across a large number of branches in two different divisions within the same firm.⁴

Another advantage of our setting is that the firm initially used profit-related pay in all divisions. This means that we are able to distinguish the impact of the particular form of incentive scheme from the general impact of any form of performance-related pay.

² See, *inter alia*, Nørreklit (2000, 2003), Jensen (2001), Cools and van Praag (2003), Gosling (2003), and Fink (2004). There is also an earlier literature; Ridgway (1956), e.g., criticizes the idea of multidimensional incentives schemes.

³ Hoque and James (2000) survey 66 Australian manufacturing firms; Banker, Potter and Srinivsan (2000) look at 18 hotels; Malina and Selto (2001) consider multiple divisions of a large firm; Ittner, Larcker, and Meyer (2003) conduct a case study in a large firm; Burgess et al. (2004) evaluate a randomized trial incentive scheme in job centers in the United Kingdom; Neely, Kennerley, and Martinez (2004) consider 35 branches in one firm; Davis and Albright (2004) look at nine branches of a firm.

⁴ Burgess et al. (2004) were able to evaluate the introduction of a team-based multitask incentive scheme in job centers in the United Kingdom, where implementation was randomized.

We find that behavior changed in response to the Balanced Scorecard but that overall this change did not lead to increased trading profits because, while sales increased, costs also increased, and in some branches the increase in costs was greater. However, the impact varied across branches. We undertook interviews with a number of individual managers to gain a better understanding of why this heterogeneity in response arose. These interviews highlight the complexity of the Balanced Scorecard and offer the explanation that experienced managers were better able to interpret the large number of indicators than were less experienced managers, and thus they were able to effectively allocate effort within the branch. We find empirical support for this.

The idea is that it is not only the incentives that matter but also the ability of managers and workers to respond to them. When it is necessary for managers and workers to perform a large number of tasks (e.g., the tasks involved in running a retail establishment), it is important that the manager be able to effectively decide where best to put both his and his workers' marginal effort. The Balanced Scorecard gives the manager additional information on past performance, but it does not tell the manager where additional marginal effort will be most effective. It gives the same incentive to all measures. This has been one of the main features that has been criticized regarding the Balanced Scorecard (Jensen 2001): it does not tell managers where marginal effort will be the most effective in improving performance. We interpret the fact that the more experienced managers were able to achieve better performance under the Balanced Scorecard as suggesting that it requires additional ability (acquired through experience) for the manager to know where increased effort will yield the greatest payoff.

The structure of the rest of this article is as follows. In Section II, we describe the firm, the Balanced Scorecard, how it was implemented, and the impact we would expect to find. Section III describes the data and our empirical approach. Section IV presents the results. Section V summarizes and concludes.

II. The Setting

We describe the firm, the incentive scheme, and how it was designed and implemented. We then discuss what impact we expect to see on performance.

A. The Firm

The firm is a multinational distributor of heating and plumbing products. It has thousands of branches in over 10 European and North American countries and employs around 50,000 people. In the United Kingdom there are four main divisions. Both divisions have several subdivisions,

or brands. We use data on the largest brand in the two largest divisions. Each brand is made up of a large number of branches. During the period we consider, branches dealt primarily with one brand and acted as relatively small trading units, employing between two and 32 staff (with a mean of 10).

We compare performance in the dominant brand in Division 1 (where the Balanced Scorecard was introduced) with the dominant brand in Division 2. Prior to August 2002, both divisions used the same incentive scheme, which was based only on branch profits. Division 2 kept this scheme after August 2002. The two divisions are similar in terms of average sales and profits per branch. They differ in that they sell distinct products, but both sell products that are used in the construction industry. Does a Division 2 branch represent a good control group for a Division 1 branch? The goods that these firms sell are bought by the same customers—builders who buy products from Division 1 will typically also buy products from Division 2—so the branches will experience similar demand shocks. In fact, since undertaking this analysis, the firm has started selling both sets of products in the same store. The branches operate in similar economic conditions; for example, they employ similar people from the same labor markets.

The largest part of firm profit comes from individual branch profits, though orders are also received at head office and head office also receives some volume discounts from suppliers. In this study, we focus on branch profits. Profits of an individual branch are the revenue earned on the sale of each product, minus the costs of sale, minus central branch costs. The main elements of cost at the branch level are the cost of goods sold, labor costs, infrastructure (including capital costs), distribution, and transport costs. Other costs include general and administration costs, spending on information technology, local marketing, advertising, and other branch-level administrative costs. These are described further in the data section below.

Effort of branch staff and branch managers can affect outcomes in the following ways:

- i) The price paid for a product can vary with each transaction (branch staff have discretion to negotiate individual prices, starting from a base price).
- ii) Staff deal directly with customers and thus have influence on the quantity, type and range of products sold.
- iii) Branch (as well as regional) managers influence the quantity sold through setting base price levels, marketing, and setting special offers.
- iv) Staff and branch (as well as regional) managers' actions can affect hiring and firing costs (through staff retention levels), volume dis-

counts (through the type and quantity of goods sold), and various other branch-level costs in a number of ways.

A typical branch is managed by a branch manager.⁵ It has an administrator who works in the office, two or three sales staff, a driver, and three or four people who work in the warehouse or stockyard. Further details are given in table 4. The role of the manager, among other things, is to decide on hiring decisions (in conjunction with head office), allocate staff to tasks, decide on special offers, and decide on the level and type of local marketing activity.

B. Performance-Related Pay in the Firm

Prior to August 2002, employees in both divisions received a bonus that was a function of branch-level profits. The bonus was allocated to branches based on a percentage of branch profits and was allocated to branch staff by the branch manager at his discretion and in consultation with the regional manager.

Senior management became concerned that this scheme was leading to dysfunctional behaviors, as has been emphasized in the literature.⁶ For example, managers had incentives to adjust the timing of capital investment where they would affect bonus payments, and they faced incentives to compete for business with other local branches because profits were calculated on a point-of-sale basis.⁷ The firm was also concerned that branch staff were not putting sufficient effort into activities that enhanced long-run profitability, such as maintaining customer loyalty and relationships with suppliers. Senior management decided to change the basis of the performance measurement and incentive scheme in the organization.

The senior management decided to perform a trial of a new scheme based on the Balanced Scorecard in Division 1 prior to rolling out that scheme across the firm. The Balanced Scorecard was designed to overcome the dual problems of subjective performance evaluation (which can give rise to various forms of bias and encourage workers to waste effort to curry favor with managers) and focusing on one key performance target (which can lead to dysfunctional behavior with workers focusing all their energy on tasks that are rewarded and ignoring those that are not included in the incentive scheme).⁸

The board devoted significant time discussing what were the objectives

⁵ The branch manager reports to a regional manager, who reports to a brand operations director, who reports to a brand managing director, who reports to a divisional managing director, who reports to the board.

⁶ See, *inter alia*, reviews by Hayes and Abernathy (1980) and Prendergast (1999).

⁷ See, *inter alia*, Oyer (2004) and Larkin (2006).

⁸ See Argyris (1952), Ridgway (1956), Holmstrom and Milgrom (1991), and Kerr (1995). For a recent review of the literature, see Prendergast (1999).

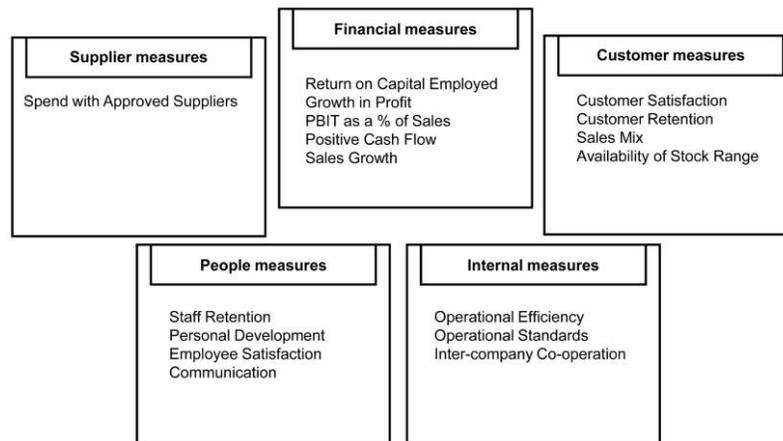


FIG. 1.—The scorecard measures

for the division, what behaviors they wanted to encourage in the branch network, and how these behaviors might be reinforced through the choice and design of appropriate performance measures. At various stages during the process, the directors consulted regional and branch managers. The firm invested considerably in education and training, as well in engagement with workers at all levels throughout this time.⁹ In the end, the firm adopted 16 measures upon which to base the new performance pay scheme. These are shown in figure 1 and further described in table 1.¹⁰

Following a 9-month design and deployment process, the Balanced Scorecard was introduced in August 2002. The scheme remained in operation until July 2004.

The new scheme made three big changes: (i) it used multiple targets rather than a single target; (ii) it was nondiscretionary rather than discretionary, and (iii) it provided more information to managers on a wide range of indicators, such as customer behavior and efficiency, than was previously available.

The specific scheme that the firm implemented worked as follows. Performance of each branch was evaluated on 16 key indicators, including both financial and nonfinancial indicators of performance. Branch staff and branch managers received a bonus based on the total number of points

⁹ The discussions were about what the objectives of the division were and who had control over them. Meetings were held in every branch, and discussion was in depth and focused on what individual workers could do to improve performance.

¹⁰ Personal development was included as a measure, but the data were not adequately collected, so they were not used. The employee satisfaction survey had a low response rate (e.g., 32% in December 2002), but the firm used it anyway.

Table 1
Balanced Scorecard Measures

Measures	Definition
Financial:	
Return on capital employed	(Profits before interest and tax) / (Debtors + Stock + Fixed Assets)
Growth in profit	Growth in profits compared to same time last year
PBIT (profit before interest and tax) as a % of sales	Profits over sales (year to date)
Positive cash flow	Cash flow over sales
Sales growth	Growth in sales compared to same time last year
Customer:	
Customer satisfaction	Score achieved via an external survey
Customer retention	Growth in number of customers retained (based on rolling 12-month measure)
Sales mix	Growth in sales of selected products compared to same time last year
Availability of stock range	Proportion of trading days where stock ins for minimum basic range are equal to or greater than 90%
Internal measures:	
Operational efficiency	Index of branch performance assessed by regional manager on four individual branch targets relating to stock, debtors, labor, and transport
Operational standards	Score from operational standards check list
Intercompany cooperation	Growth in number of customers referred to other branches compared to same time last year
People:	
Staff retention	Number of voluntary leavers as a proportion of average head count (based on a rolling 12 months)
Employee satisfaction	The number of people who indicate they are satisfied at work as a proportion of the average number of employees over the period
Communication	The number of people who feel they have been made aware of businesses activities as a proportion of the average number of employees over the period (measured by region)
Supplier:	
Spend with approved suppliers	Purchases from preferred suppliers as a share of all purchases from suppliers

that the branch earned in each month times the value of a point.¹¹ Points were calculated and are reported on a monthly basis, but they were paid on a 6-month basis.

¹¹ We only consider branch profits here. For regional managers, the bonus was based on the average of the points earned in branches under their control. For central office staff, it was calculated based on a simplified version of the Balanced Scorecard containing only measures that the central office staff could affect.

Points were calculated as follows. For each of the 16 measures, each branch is graded “green,” “amber,” or “red.” A green grade earns 3 points, an amber grade 1.5 points, and a red grade 0 points. There are 16 measures, so the maximum number of points a branch can earn in a month is 51.

The 16 measures are shown in table 1. For each measure, the firm specified thresholds for obtaining a green, amber, or red grade. For example, consider the first measure, return on capital employed. This is measured by profit before interest and tax (PBIT) over capital employed (debtors plus stocks plus fixed assets). A value of 25% or more earned a branch a green grade, a value of between 20% and 25% earned an amber grade, and a value of less than 20% earned a red grade. Similar thresholds were defined for each of the other measures. Table 1 shows that the measures are a combination of absolute targets that the firm wants all branches to meet and a few targets that require branches to continually improve on their own past performance (there are four measures that fall into this latter category—growth in profit, sales growth, customer retention, and intercompany cooperation).

In 2003 a point was worth £1 for branch staff (the value of a point is higher for managers), so the maximum bonus a branch worker could earn in a month was £51, or £612 a year. Salaries for branch workers average around £12,000, so the maximum bonus represents around 5.1% of salary.

The total amount the firm allocated for the scheme was just under £2 million. This was set aside in a separate account to signal the firm’s commitment to the scheme. The previous scheme, based on profits, cost about the same total amount.¹² As well as changing the incentive structure, an important aspect of the Balanced Scorecard was the provision of more detailed information on performance, and, crucially, on several leading indicators of profitability. Each branch manager received a detailed report every month on the performance of his or her branch on each of the 16 measures.

Why did the firm implement the Balanced Scorecard in Division 1 and not Division 2? This is an important question as our strategy for identifying the impact of the Balanced Scorecard relies on performance being independent of this decision. The pragmatic reason is simply that the divisions were run relatively independently and the managing director of Division 1 was rather more keen on management innovation, particularly on the Balanced Scorecard. At the outset, the organization considered whether it would subsequently implement the Balanced Scorecard in other divisions, but, in essence, it decided to delay a decision and instead treat Division 1’s implementation as a pilot study. Thus, we argue that the

¹² The firm estimates the cost of implementation at around £0.5 million, including the direct and indirect costs, such as management time.

introduction can be treated as independent of any expectations about the impact the Balanced Scorecard would have in one division over the other.

In evaluating the impact of the Balanced Scorecard, we need to be careful to distinguish two questions: (i) Did the new incentive scheme change behavior? (ii) Did this behavior lead to improved performance? The idea behind the Balanced Scorecard is that the firm needs to determine what are the key drivers of future profitability and build these into the scheme.

C. Anticipated Impact on Performance

How do we expect the key changes—(i) introduction of multiple measures, (ii) move from discretionary to nondiscretionary, (iii) move from individual to group evaluation of performance, and (iv) increased information—to affect performance? Remember that, as mentioned above, there was no major change in the overall value of performance-related pay, only in its form.

In order to consider how this affects performance, we consider the incentives that individuals faced and how they changed. We assume that individuals seek to maximize their income, net of effort. Income consists of a base wage and a performance-related payment. Performance is a function of the effort of all staff, managers, and directors. Effort is costly. Individuals will exert effort up to the point where the marginal cost of effort equals the marginal benefit in terms of the incentive payment. This is true under either scheme. What changed is the way performance was measured and rewarded.

The main objectives of the firm were to encourage workers to put more effort into a broad range of activities that were previously not rewarded but which the directors of the firm believed feed into firm-level long-term profits (long term being over several months).¹³ The tasks concerned are substitutes from the individual worker's point of view (they each take time and an individual worker can spend time on one task or another), but, from the point of view of the firm and value maximization, they are complements, in the sense that workers need to spend time on all of the tasks in order to maximize the value of the firm. For example, a worker can either sweep the floor or restock the shelves—from the worker's point of view, these are substitute activities. But from a customer's point of view, these are complementary characteristics of a shop—they want both a clean shop and well-stocked shelves.

Where workers undertake tasks that are substitutes for each other and

¹³ This increase in firm performance could arise, e.g., if the activities of workers in one branch affected the performance of another branch or if the activities of workers in one period affected the performance of the branch in future periods.

where the rewards are equal and independent (as is the case here), workers will devote more time to those tasks they find easiest to perform. If the measurement of some tasks is more precise or more clearly understood—in the sense that it is easier to identify improvements in performance—then workers will devote more time to those tasks that are measured more precisely.¹⁴ This is essentially the objection that Jensen (2001) and others have made to performance-related pay schemes like the Balanced Scorecard—it does not give clear guidance to workers as to what is the most important task to perform (should they spend another 10 minutes sweeping the floor or should they start stocking the shelves), but rather it allows them to put too much effort into unprofitable tasks (either because they have misinformation about the payoff to the individual tasks or because both are rewarded, so they simply do the one that is least costly).

The Balanced Scorecard scheme with 16 measures was complex, and managers were provided with large amounts of information. In order to effectively use this information, the manager needed to be able to assimilate it and understand what it meant in terms of future performance. A key issue is that the data presented in the Balanced Scorecard did not give managers the information they needed to take action. Each manager needed to be able to translate the information provided on the Balanced Scorecard so that it told them something specific about the local situation. For example, consider one of the measures—customer retention. Each month the manager would be given information on how many customers they were retaining compared to the same time last year. This is a useful leading indicator of profitability. However, simply knowing that you retained 5% fewer customers than this time last year is not sufficient information to enable the manager or shop workers to act. They needed to be able to figure out which customers were not coming back and why in order to know where to put effort that would increase retention rates. This is where experience comes in. Our thesis, developed through interviews with branch managers, is that experienced managers are either more likely to have seen in the past, or are better able now to develop, local solutions that allow them to take action to improve performance according to the Balanced Scorecard. A more experienced manager will be better able both to interpret the large number of indicators and to (e.g., more credibly) motivate staff to put effort into the activities that matter for performance; with an inexperienced manager, workers can easily become overloaded with information and tasks, leading to underperformance. We explore this idea in our empirical analysis below.

¹⁴ See Holmstrom and Milgrom (1991) and Burgess et al. (2004).

III. Data and Econometric Method

A. Data

The main data are drawn from the monthly profit and loss accounts of the firm. We have information at the branch level on sales, gross profits (sales minus cost of goods sold), trading profits, labor costs, infrastructure, transport costs, general and administrative costs, information technology (IT) investment, local marketing and advertising expenditure, and other costs. We have these data monthly from August 1999 to July 2005. We also have data on each employee in each branch in Division 1, including age, job title, and length of tenure in the firm from August 2003 to July 2004.

We use the location (postcode) of each branch to match branches from the two divisions. There are a number of factors that affect sales but that are both exogenous (not affected by actions of the firm) and will have a common effect across both divisions. These include the economic cycle, local economic and labor market conditions, and other local factors. There may, however, be some variation in demand conditions. The products sold by Division 1 are used both inside and outside buildings, while those sold by Division 2 are predominantly used inside. Thus weather conditions will have a differential effect on demand for the two goods. To control for this, we use monthly data from 26 weather stations in the United Kingdom on the minimum temperature (in Celsius) and rainfall (in millimeters). In addition, goods sold by Division 2 are used more intensively in refitting houses, while goods sold in Division 1 are used more intensively in new buildings. To control for this, we use data on total quarterly construction activity in a range of categories to capture variation in aggregate demand. These data come from the Construction Products Association. We aggregate separate information on aggregate quarterly output for new private sector housing, private and public sector remodeling, and commercial and industrial development. The firm has given us the weight of each of these in demand for each of their brands, and we use this to construct a measure of quarterly aggregate demand for each brand.

Tables 2 and 3 show descriptive statistics of the main variables. Table 4 shows descriptive statistics on the number of employees, types of jobs within a branch, and the average experience and age of staff.

We match each branch from Division 1 Brand A to the geographically nearest branch from Division 2 Brand C. The markets for the goods sold by the firm are very local—the firm tells us that consumers are rarely willing to travel further than 30 kilometers. There are four Division 1 branches that do not have a Division 2 branch within 30 kilometers, and we exclude these from our analysis. The average distance between matched branches is 4.5 kilometers.

Table 2
Means (in £ Thousands) for Division 1 Branches and Matched
Division 2 Branches

	Division 1	Division 2 (Matched Only)
Branches	156	121
Observations	11,076	11,076
Sales	152.5 (87.8)	166.4 (103.4)
Gross profits	45.7 (28.3)	39.9 (19.1)
Trading profits	15.3 (19.4)	22.8 (15.6)
Labor costs	14.9 (7.7)	8.8 (4.2)
Infrastructure	7.6 (4.7)	4.5 (2.4)
Transport costs	4.2 (2.6)	1.1 (1.1)
General and administration	1.4 (1.3)	.8 (.6)
Other	.9 (.7)	.7 (.5)
IT	.48 (.22)	.49 (.13)
Marketing and advertising	.15 (.34)	.02 (.15)

NOTE.—Values are monthly in nominal £ thousands over the period August 1999 to July 2005. Numbers in parentheses are standard deviations.

B. Econometric Method

Our setting allows us to adopt a quasi-experimental design method and use a combined matching and difference-in-difference estimator.¹⁵ As highlighted above, a major problem in the literature that attempts to identify the impact of incentive schemes has been the fact that organizations choose whether and which incentive schemes to adopt—the adoption of the incentives scheme is endogenous.

We are able to use the fact that the firm implemented the Balanced Scorecard in one part of the firm but not the other to tackle this problem. Differencing between the matched branches allows us to control for all of those factors that have a similar effect on the two branches. However, the firm has highlighted two specific factors that may affect demand for the products sold in the two divisions differently, and we control for these in our analysis. We believe that this intrafirm interdivision comparison gives us a substantially better control group than is generally used in the literature (where, e.g., a different firm is generally used, where the differences between treatment and control will be much greater).

¹⁵ See, inter alia, Cook and Campbell (1979) and Heckman, Ichimura, and Todd (1997). For a recent survey, see Blundell and Costa-Dias (2000, 2005).

Table 3
Descriptive Statistics—Weather and Demand

	Mean
Minimum temperature (in Celsius), measured at 26 points throughout the United Kingdom	7.28 (4.17)
Rainfall (in millimeters)	65.7 (41.0)
National quarterly demand for activities using Division 1 products (in £ millions)	2,628 (200)
National quarterly demand for activities using Division 2 products (in £ millions)	2,588 (168)

SOURCES.—Data on weather are from <http://www.met-office.gov.uk/climate/uk/stationdata/>. Data on quarterly demand were provided by the firm and are from industry sources.

NOTE.—Standard deviations are in parentheses.

To formalize this, we can write the determinants of sales for each type of branch,

$$S_{irt} = \alpha_i + \beta_1 \eta_{rt} + \lambda_1 D_{1t} + \phi_1 \tau_t + \gamma_1 W_{rt} + \delta_i \text{BSC} + e_{irt}, \quad (1)$$

$$S_{jrt} = \alpha_j + \beta_2 \eta_{rt} + \lambda_2 D_{2t} + \phi_2 \tau_t + \gamma_2 W_{rt} + \varepsilon_{jrt}, \quad (2)$$

where branches within Division 1 are indexed i , branches within Division 2 are indexed j , and r indexes regions and t time (months). The variables that we observe are S , sales; D , national demand; W , weather; and BSC, use of the Balanced Scorecard. The time subscript on the coefficient indicates that we allow the impact of the Balanced Scorecard to vary over time. This makes the specification flexible; for example, it could capture anticipation effects that occur before the Balanced Scorecard was put in place or it could pick up effects that last beyond the time period that the Balanced Scorecard was in operation. We explore these issues below.

The variables that we do not observe include α_i and α_j , which are branch fixed effects that capture unobserved time-invariant branch characteristics; η_{rt} , which captures unobserved factors at the regional level that might vary over time, such as variation in labor market conditions, supply chain issues, or factors that affect local levels of demand; τ , which is a set of common seasonal (monthly) effects; and e and ε , which are idiosyncratic shocks.

Omitting the branch fixed effects could lead to potential bias in the estimated effect of the Balanced Scorecard if the adoption of the Balanced Scorecard was correlated with other unobserved branch characteristics, such as manager quality. However, these branch dummies are time invariant and so will not control for any time-varying factors that are correlated with BSC, for example, local labor market conditions or supply chain issues that might affect a manager's ability to respond to the Balanced Scorecard. If we are willing to assume that these local market con-

Table 4
Job Type, Number of Employees, and Average Experience

Job	Number of Workers per Branch				Mean Experience	Mean Age
	Mean	Standard Deviation	Minimum	Maximum		
Manager	.883	.359	0	2	12.97 (9.88)	42.09 (8.57)
Office	1.090	.808	0	3	8.82 (9.06)	41.24 (11.83)
Sales	2.548	1.667	0	9	7.09 (8.67)	38.53 (12.80)
Driver	1.651	1.209	0	6	6.52 (7.75)	46.44 (10.58)
Warehouse and stockyard	3.670	2.107	0	14	5.77 (7.67)	38.97 (13.33)
Other	.354	.543	0	2	7.58 (7.48)	47.57 (13.56)
Total	10.2	4.631	3	30	7.23 (8.52)	40.88 (12.64)

NOTE.—*Manager*: branch manager, branch manager (designate). *Office*: administrative assistant, administration supervisor, administrator, assistant branch manager, assistant depot manager, branch supervisor, deputy manager, estimator, general clerk, typist, office manager, secretary. *Sales*: contracts/sales administrator, credit controller, goods inwards assistant, inside sales, sales supervisor, sales clerk, sales assistant, sales negotiator, sales representative, showroom supervisor, stock controller. *Driver*: driver, glass cutter/driver, warehouse assistant/driver. *Warehouse/Yard*: depot manager, depot manager designate, drainage supervisor, foreman, heavy supervisor, hire assistant, ironmongery supervisor, logistics manager, maintenance supervisor, operations manager, product supervisor, stores assistant, warehouse assistant, yard assistant, Saturday assistant, timber supervisor, transport supervisor, warehouse supervisor, yard manager, yard supervisor. *Other*: fixer, machine operator, mill operative, cleaner, timber machinist, trainee.

ditions affect the two types of branches in the same way (as argued above), then $\beta_1 = \beta_2$. If we match each branch from Division 1 to the geographically nearest branch from Division 2 and consider the difference between the two branches, then these unobserved region specific characteristics difference out.

We further assume that changes in aggregate demand feed through into sales and profits in the same way across the two branches (so that $\lambda_1 = \lambda_2$), which means that we can include the difference in aggregate demand for goods sold in the two branches. This gives us:

$$(S_{irt} - S_{jrt}) = (\alpha_i - \alpha_j) + \lambda(D_{1t} - D_{2t}) + (\phi_1 - \phi_2)\tau_t + (\gamma_1 - \gamma_2)W_{rt} + \delta_t \text{BSC} + (e_{irt} - e_{jrt}), \quad (3)$$

which is a combined matched and difference-in-differences estimator.

To investigate the idea that there was a heterogeneous impact of the introduction of the Balanced Scorecard, we extend this specification to allow the impact of the Balanced Scorecard to vary with other observable characteristics of the branches, in particular, the level of experience of senior staff; that is, we replace δ_t with $\delta_{1t} + \delta_{2t} \text{Experience}$.

Table 5
Matched/Difference-in-Difference Results on Sales, Different Time Periods

	(1)	(2)	(3)	(4)
November 2001 to July 2002				404 (2,382)
BSC, August 2002 to July 2004	4,538 (1,786)**			
BSC, August 2002 to July 2003		3,752 (2,624)	8,305 (2,435)***	8,671 (3,577)**
BSC, August 2003 to July 2004		4,992 (2,136)**	8,903 (4,020)***	9,216 (4,625)**
August 2004 to July 2005			6,916 (4,872)	7,124 (5,157)
Demand	-25 (28)	-23 (30)	-50 (24)**	-52 (20)***
Minimum temperature in Celsius	1,686 (455)***	1,694 (463)***	1,354 (453)***	1,348 (455)***
Rainfall in millimeters	-90 (16)***	-89 (15)***	-80 (13)***	-79 (13)***
R^2	.15	.15	.15	.15

NOTE.— $N = 11,076$ observations on 156 branches over the period August 1999 to July 2005. Constant and month and branch dummies are included in all regressions. Dependent variable is the difference in sales in a Division 1 branch and the geographically nearest Division 2 branch. Demand is the difference in national demand for Division 1 products and demand for Division 2 products. BSC = Balanced Scoreboard in use.

** Significant at the 5% level.

*** Significant at the 1% level.

IV. Results

A. The Overall Impact of the Balanced Scorecard

We start in table 5 by considering how we parameterize the BSC variables. The dependent variable is the difference in sales between the matched branches. In the first column, we include an indicator for just the 2 years that the incentive scheme was in place (August 2002 to July 2004), so we are comparing the level of sales in Division 1 branches with Division 2 branches during this period with the difference in the level before and after this period. This shows that sales in Division 1 branches increased by an average of £4,538 per month more than Division 2 branches during the 2-year period in which the Balanced Scorecard was in place. In column 2, we split this period in half to see if the impact was similar across the 2 years. The reason for doing this is that we would expect to see year-on-year increases if the scheme was successful. What we see is that there is a significant increase in sales in year 1, a significant and larger increase in year 2, and that, when use of the Balanced Scorecard as a performance-related pay scheme is phased out, some of the gains are sustained.¹⁶ In column 3, we also include an indicator for the year after the firm put the incentive scheme on hold—so we are now comparing

¹⁶ As discussed below and shown in fig. 6, this is in large part because the variance of performance increases across branches.

only to the 2 years before the Balanced Scorecard was implemented. We now see a larger impact from the Balanced Scorecard, with a decline after the scheme was put on hold, though perhaps not immediately back to the original levels. In column 4, we consider whether there was an anticipation effect in the initial discussion period,¹⁷ but we find no evidence of this. For the remainder of this article, we stick to the parameterization in column 3. In table 5, we also see that demand for Division 2 products grew somewhat faster than that for Division 1 products (the coefficient on demand is negative) and that there is more demand for Division 1 products when it is hotter and less when it is raining.

These results suggest that the Balanced Scorecard had an impact through increasing sales, but what about profits? In table 6, we look at how gross profits, trading profits, labor costs, infrastructure expenditure, transport costs, general and administrative costs, information technology expenditure, local marketing, and advertising expenditure were affected.

The first column of table 6 repeats column 3 of table 5. All regressions include controls for aggregated demand, the weather, month, and branch effects. In column 2, we see that there was also a corresponding increase in gross profits (the difference between sales and gross profits is the cost of goods sold). In column 3, we see that this did not feed through into an increase in trading profits and that in fact this led to a fall in profits in the later period of, on average, $-\text{£}1,018$. The difference between gross and trading profits is the costs considered in columns 5–9. The fall in profits arose largely because transport costs (col. 6) and labor costs increased (col. 4) and, to a lesser extent, because infrastructure (col. 5) and general and administrative costs (col. 7) increased. The IT costs (and Other, not shown) actually fell.

This is the first empirical result of this article: the Balanced Scorecard led to an increase in sales, but costs increased by at least as much as sales at the aggregated branch level, so there was no increase (and actually some decrease) in profits. We emphasize these results in figures 2 and 3.

To produce figure 2 we ran a regression similar to that shown in column 1 of table 6 but with a separate time dummy for each month.¹⁸ Figure 2 shows a plot of the monthly dummies smoothed over the period shown. The figure shows that, prior to the introduction of the Balanced Scorecard in August 2002 (200208), the difference in sales between the Division 1 and Division 2 branches was fairly constant. After the introduction of the Balanced Scorecard sales in Division 1, those branches grew more

¹⁷ This is what is often called a Hawthorne effect in anticipation of the actual implementation.

¹⁸ That is, we ran a regression of the difference in the level of sales between each Division 1 branch and its matched Division 2 branch regressed on branch fixed effects, seasonal fixed effects, weather, and time dummies.

Table 6
Matched/Difference-in-Difference Results on All Variables

	Sales (1)	Gross Profits (2)	Trading Profits (3)	Labor Costs (4)	Infrastruc- ture (5)	Transport Cost (6)	General and Administration (7)	IT (8)	Marketing and Advertising (9)
BSC: August 2002 to July 2003	8,305 (2,435)**	1,512 (591)**	771 (677)	192 (161)	91 (125)	704 (82)**	28 (45)	-59 (8)**	12 (14)
BSC: August 2003 to July 2004	8,903 (4,020)**	2,565 (976)**	995 (846)	675 (261)**	183 (211)	1,014 (112)**	115 (49)**	-78 (10)**	19 (18)
August 2004 to July 2005	6,916 (4,872)	1,835 (1,155)	-1,018 (932)	933 (299)**	667 (309)**	1,154 (119)**	88 (50)*	-91 (10)**	-4 (10)
Demand	-49 (23.60)**	-4.08 (5.33)	-7.61 (5.09)	-2.161 (1.43)	1.732 (1.01)*	-594 (-0.6)	.200 (.33)	-.333 (.07)**	-.163 (.09)*
Minimum temperature in Celsius	1,354 (452.77)**	527 (1,35.38)**	624 (157.24)**	-6.60 (30.43)	-88.9 (55.78)	-40.6 (18.04)**	34.9 (10.62)**	-5.457 (2.42)**	372 (2.58)
Rainfall in millimeters	-79 (13.35)**	-19.0 (3.39)**	-22.8 (4.11)**	.144 (.74)	-.957 (.96)	.167 (.45)	-.724 (.38)*	.017 (.05)	.100 (.10)
R ²	.15	.20	.13	.05	.01	.08	.01	.05	.00

NOTE.—Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Standard errors in parentheses are clustered at the branch level. N = 11,076 observations on 156 branches over the period August 1999 to July 2005. Constant and month and branch dummies are included in all regressions. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

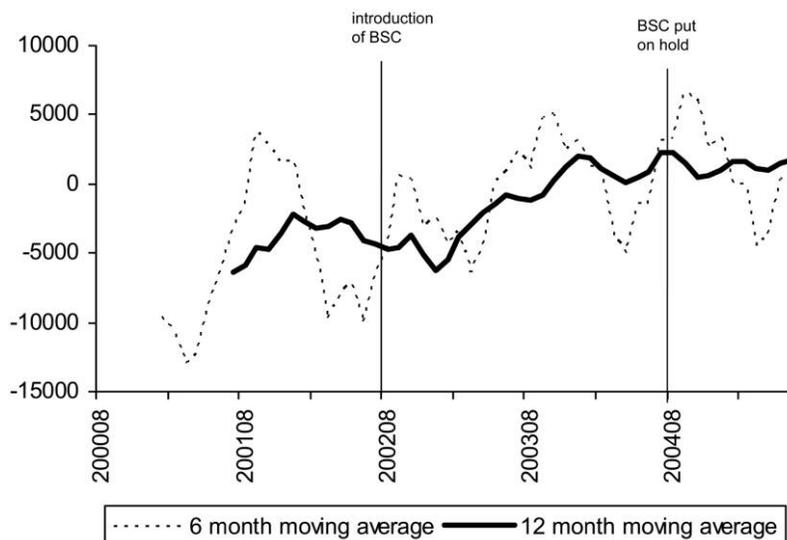


FIG. 2.—Sales in all Division 1 branches relative to Division 2 branches. Each line shows the moving average of monthly time dummies estimated from a regression that includes demand, weather, and branch dummies (so similar to that contained in col. 1 of table 6), except where the Balanced Scorecard and monthly dummies have been replaced by 59 individual month dummies that run from August 2008 to the end of the period. The data include 11,076 observations on 156 branches over the period August 1999 to July 2005. The figure shows the 6-month and 12-month moving average of the 59 estimated coefficients.

rapidly. When the Balanced Scorecard was put on hold, the difference in the level of sales evened out again.

To produce figure 3 we ran a similar regression but now with the difference in trading profits as the dependent variable. The figure plots the monthly dummies smoothed over the period shown. The figure shows that, prior to the introduction of the Balanced Scorecard in August 2002, the difference in trading profits between the Division 1 and Division 2 Branches was declining. After the introduction of the Balanced Scorecard, trading profits in Division 1 stopped falling, but they did not grow in the same way that sales did (this is because a number of large components of costs grew more rapidly in Division 1 branches than in Division 2 branches). When the Balanced Scorecard was put on hold, the difference in the level of trading profits did not change.

B. Did the Balanced Scorecard Target the Right Tasks?

One important question is whether the Balanced Scorecard targeted the right tasks, that is, tasks that would improve performance. Did branches that improved on the nonfinancial aspects measured subsequently experience improved performance? Did the Balanced Scorecard correctly iden-

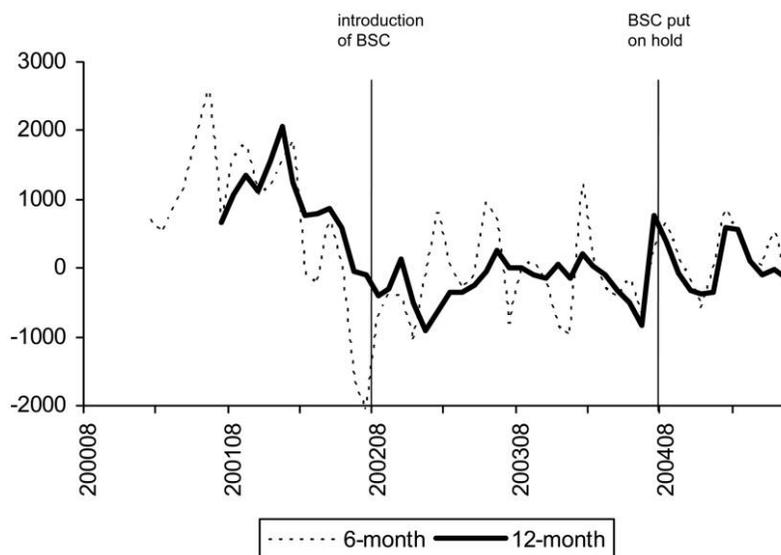


FIG. 3.—Trading profits in all Division 1 branches relative to Division 2 branches. Each line shows the moving average of monthly time dummies estimated from a regression that includes demand, weather, and branch dummies (so similar to that contained in col. 1 of table 6), except where the BSC and monthly dummies have been replaced by 59 individual month dummies that run from August 2008 to the end of the period. The data include 11,076 observations on 156 branches over the period August 1999 to July 2005. The figure shows the 6-month and 12-month moving average of the 59 estimated coefficients.

tify the key drivers of performance? We use the data collected under the Balanced Scorecard to distinguish between branches that successfully put effort into improving on the nonfinancial measures from those that did not. For each branch, we calculate what proportion of the total nonfinancial points that were available they earned. This ranges from 16% to 74% and has a median value of 52%.

In table 7, we split the sample into those branches that were below and above the median share of points earned on the nonfinancial measures. In the left-hand side of table 7, we consider the change in outcomes of those branches that either did not try—or were not able—to perform well on the nonfinancial measures, and on the right-hand side, those branches that did well on the nonfinancial measures. We focus on sales and trading profits.

We see large differences between the two groups. In the left-hand panel, branches that did poorly on nonfinancial measures did not experience any significant growth in sales, but they did experience some increases in costs, leading to reduced trading profits. In contrast, those branches that did well on the nonfinancial measures saw an increase in sales and trading

Table 7
Matched/Difference-in-Difference Results on All Variables: Comparison of
Branches on Nonfinancial Balanced Scorecard Measures

	Branch Score on Nonfinancial Balanced Scorecard below Median		Branch Score on Nonfinancial Balanced Scorecard above Median	
	Sales (1)	Trading Profits (2)	Sales (3)	Trading Profits (4)
BSC: August 2002 to July 2003	1,947 (3,740)	-838 (1,062)	14,676 (2,964)***	475 (167)***
BSC: August 2003 to July 2004	3,535 (6,135)	-979 (1,260)	14,266 (5,173)***	822 (230)***
August 2004 to July 2005	3 (7,538)	-2,725 (1,437)*	13,583 (6,097)**	1,182 (278)***
Demand	-44.59 (31.829)	-7.457 (7.277)	-55.526 (35.113)	-2.122 (-2.133)
Minimum temperature in Celsius	1,063.194 (671.935)	631.729 (216.424)***	1,676.383 (568.785)***	-7.446 (-38.592)
Rainfall in millimeters	-48.264 (18.298)**	-19.621 (5.172)***	-113.104 (18.036)***	-.652 (-.977)
R^2	.12	.09	.21	.08

NOTE.—Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Standard errors (in parentheses) are clustered at the branch level. Each regression contains 5,538 observations on 78 branches (because the sample is split in half at the median) over the period August 1999 to July 2005. Constant, month and branch dummies, demand, minimum temperature, and monthly rainfall are included in all regressions. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

profits. While costs did increase in these branches, sales increased by more, resulting in higher profits. Figure 4 shows a similar picture. Each dot represents a branch. The x -axis shows the share of possible nonfinancial points earned, and the y -axis shows branch profits (scaled by sales). The correlation between these is 0.66 and is statistically significant.

These results provide some evidence that the Balanced Scorecard measures were well chosen—those branches that did well on the nonfinancial performance indicators also did well financially. But, one concern is that this could simply be showing that good branches (or good branch managers) do well on both financial and nonfinancial indicators. Proponents of the Balanced Scorecard argue that giving branches an incentive to put effort into a broader range of factors that feed into long-run performance will lead to better performance in the long run. Opponents of the Balanced Scorecard argue that giving managers and workers so many different incentives will lead to worse performance as individuals lose focus and put too much effort into the easiest tasks. To investigate this, we make a further comparison. We consider branch financial performance prior to

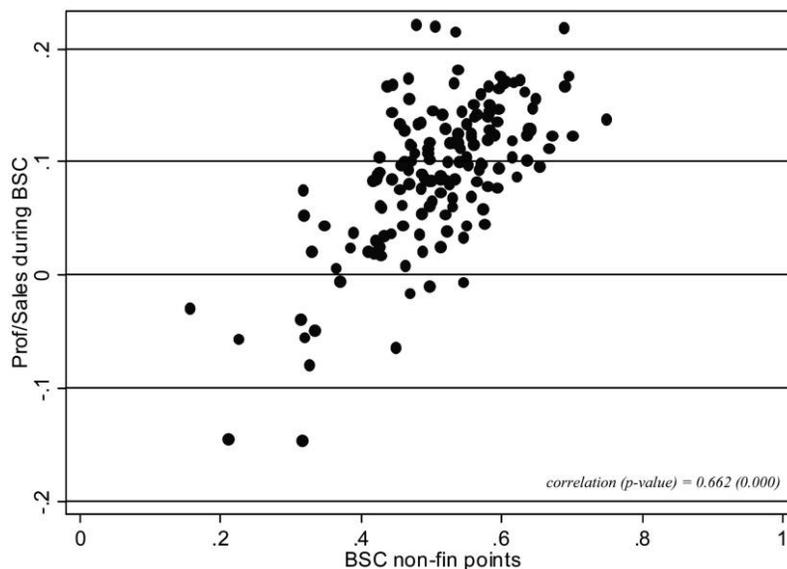


FIG. 4.—Comparison of nonfinancial with financial performance during Balanced Scorecard, using profits over sales rather than BSC measures. There are 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the nonfinancial BSC measures (customer, internal, people, and supplier) over the period August 2002 to July 2004; y-axis is the average value of trading profits over sales for each branch over the period August 2002 to July 2004.

the Balanced Scorecard (August 1999 to July 2002) and during the Balanced Scorecard (August 2002 to July 2004). We split the sample by financial performance in the period before the Balanced Scorecard was introduced.

In figure 5, we show that good performance on the nonfinancial points during the Balanced Scorecard led to good financial performance afterward in both cases, and the relationship is stronger in those branches that previously had performed badly than in those that previously had done well (correlation coefficient [p -value] of 0.727 [0.000] and 0.346 [0.002], respectively).

These pictures are suggestive, but we have not controlled for any of the local economic time-varying factors we were concerned with above. To do this, we return to the regressions and compare performance in four groups of branches: (1) those that did badly on financial measures prior to implementation of the Balanced Scorecard and who subsequently also did badly on nonfinancial measures during the implementation of the Balanced Scorecard, (2) those who did badly on financial measures prior to implementation and did well on nonfinancial during implementation, (3) those who did well on financial measures prior but badly on nonfi-

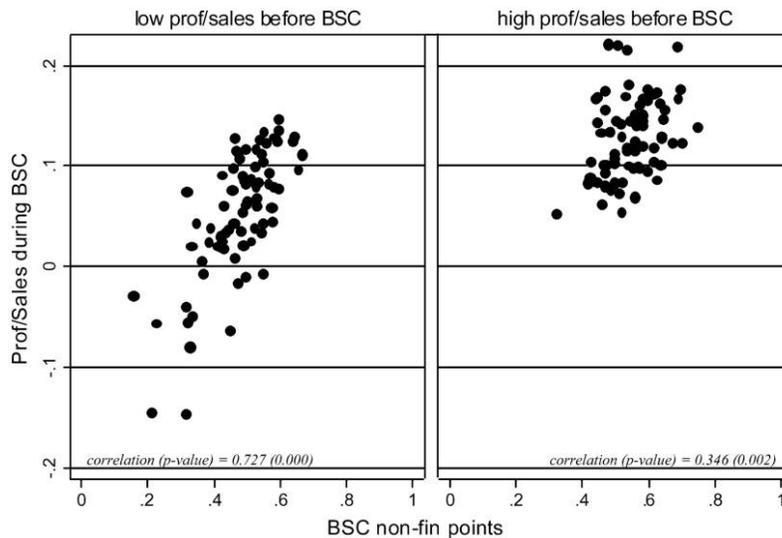


FIG. 5.—Comparison of nonfinancial with financial performance during Balanced Scorecard, split by financial performance before Balanced Scorecard. There are 156 observations (one for each Division 1 Brand A branch); x -axis is the share that each branch earned of the total points they could have earned on the nonfinancial BSC measures (customer, internal, people, and supplier) over the period August 2002 to July 2004; y -axis is the average value of trading profits over sales for each branch over the period August 2002 to July 2004. Left-hand panel shows those branches that had below-median ratio of profit to sales over the period August 1999 to July 2002.

financial measures during, and (4) those who did well on financial measures prior and well on nonfinancial measures during. We focus on the results for sales and trading profits.

The far-left-hand panel of table 8 shows the estimates for group 1, the second panel for group 2, the third panel for group 3, and the far-right-hand panel for group 4.

What we see is that branches that do well on the nonfinancial measures also do well on financial measures, regardless of whether their financial performance was above or below average before the introduction of the Balanced Scorecard. In particular, those branches that previously did badly on financial measures (second panel from the left in table 8) had substantially higher sales after implementation of the BSC and they had some increase in trading profits.

Consider the branches that did well financially prior to implementation. Those that did badly on nonfinancial measures (third panel from the left in table 8) did badly in financial terms after implementation, while those that did well on nonfinancial measures (far right panel) did well in financial terms after implementation.

Table 8
Matched/Difference-in-Difference Results on All Variables: Comparison of Branches
on Nonfinancial Balanced Scorecard Measures

	Branch Score on Financial Measures Pre-BSC below Median			Branch Score on Financial Measures Pre-BSC above Median			
	Branch Score on Nonfinancial BSC Measures below Median		Trading Profits (4)	Branch Score on Nonfinancial BSC Measures below Median		Trading Profits (6)	
	Sales (1)	Profits (2)	Sales (3)	Profits (5)	Sales (7)	Profits (8)	
BSC: August 2002 to July 2003	1,665 (5,125)	-418 (1,190)	19,833 (5,786)**	2,275 (5,396)	11,454 (3,15)**	-1,428 (2,017)	2,265 (1,070)**
BSC: August 2003 to July 2004	678 (8,076)	-147 (1,475)	20,045 (10,554)*	7,908 (9,544)	10,631 (5,238)**	-2,291 (2,293)	3,227 (1,429)**
August 2004 to July 2005	-7,080 (10,227)	-2,269 (1,837)	16,018 (10,514)	10,617 (10,827)	11,981 (7,505)	-3,601 (2,335)	1,444 (1,663)
Demand	-53,493 (33,908)	-3,43 (7,265)	-139,816 (58,944)**	-31,101 (62,969)	-3,042 (42,45)	-14,723 (14,902)	-6,414 (9,887)
Minimum temperature in Celsius	1,098,729 (876,553)	281,484 (252,039)	457,288 (928,409)	1,483,235 (1,014,43)	2,484,841 (726,715)**	1,303,259 (364,396)**	964,148 (298,059)**
Rainfall in millimeters	-19,043 (20,931)	-10,703 (5,808)*	-117,575 (32,175)**	-112,138 (36,930)**	-112,307 (22,354)**	-39,426 (11,212)**	-24,764 (7,177)**
Observations	3,408	3,408	2,130	2,130	3,408	2,130	3,408
Branches	48	48	30	30	48	30	48
R ²	.1	.06	.16	.17	.27	.2	.22

NOTE.—Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Standard errors (in parentheses) are clustered at the branch level. Sample period is August 1999 to July 2005. Constant, month and branch dummies, demand, minimum temperature, and monthly rainfall are included in all regressions. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

To summarize, these results suggest that those branches that were successful on the nonfinancial measures also did well financially. The results also suggest that there was heterogeneity in the effect of the Balanced Scorecard across branches. We now turn to investigate what might explain this heterogeneity.

C. The Importance of Experience

From a theoretical perspective, why should the impact of the Balanced Scorecard vary by the experience of managers? Some studies have argued that less experienced managers rely on formal measurement systems to supplement their inexperience more than do experienced managers (Bourne, Kennerley, and Franco 2005). Empirically, we find the opposite, that the more experienced the manager, the greater the impact of the Balanced Scorecard. As an incentive scheme, there is no obvious reason why more experienced managers should be more motivated by the incentive scheme than less experienced managers (if anything, we might think the opposite, as the incentive will account for a lower share of their total income). The consumer psychology literature makes an important distinction between motivation, ability, and opportunity (Batra and Ray 1986; Andrews 1988; MacInnis, Moorman, and Jaworski 1991), which is also discussed in the organization design and psychology literatures (Parker and Wall 1998). This literature makes explicit the distinction between *opportunity* (the freedom and scope to act), *motivation* (the incentives to respond), and *ability* (whether an individual has the necessary skills and capabilities to respond). The point made in both the consumer psychology and the organization design literatures is that the absence of any of the above factors significantly limits the likelihood either of action or the impact of action. In the case of the Balanced Scorecard, it is clear that there is an incentive for managers to act—hence, the motivation is in place—but do they have the opportunity and the ability to act? And are the answers to these questions affected by their experience?

To investigate this, we undertook 20 systematic telephone interviews with branch managers. The branch managers selected for interview had all been with the firm for at least 4 years; hence, they had been in the post for the entire period during which the Balanced Scorecard operated. Each interview lasted between 45 and 75 minutes, during which time branch managers were asked a series of questions about their experiences with the Balanced Scorecard. Initially the interviewees were asked to explain when they first heard about the Balanced Scorecard and what their initial reactions were to the scheme. Next, they were asked to explain how they introduced the Balanced Scorecard to their staff and to comment on the staff's reaction to the scheme. Third, the interviewees were asked to explain how they used the Balanced Scorecard and how they involved their staff in discussing the results. Fourth,

they were asked to comment on what happened to the Balanced Scorecard over the time it was in operation. Finally, they were asked to comment on how they measured and managed performance in their branches today (post the Balanced Scorecard).

From the interviews we found out that, while the majority of managers interviewed were positive about the scheme, several raised issues that led us to question whether they were able to act on the data. The first common theme to arise was the issue of data quality: “I think one of the problems straight away was we didn’t have the ability to measure a lot of the measures correctly or the figures were wrong” (Branch Manager A).

The second was whether the measures were under the manager’s control:

Personally, we would discuss why it had happened and we would put an action plan in, just a vehicle for action plan, what we could try and do to improve the score, but there were certain measures that you couldn’t control, like your cash collection. If a customer had no money, there was absolutely nothing you could do to encourage them to pay the bill, so that was totally out of your control. (Branch Manager B)

Quite honestly, I used to look at the control card and I used to see the reds, I used to see the yellows, I used to see the greens, and I knew which ones I could influence easily. Okay, and I could do something about those, and I would do it. The ones I couldn’t do, I just had to leave. (Branch Manager C)

Taken together, these two issues—the quality of the data and the question of whether the measures were under the managers’ control—clearly have an impact on a manager’s opportunity and ability to act, as well as on their motivation to act. Poor-quality data undermine the managers’ confidence in the scheme, causing them to question whether the scheme is having a positive effect.

Interestingly, the issue of control over the measures is a factor that many managers raised, but then some found solutions too. For example, one experienced manager explained the local work-around that he and his team had developed to enable him to act on the customer retention data that other managers said they had no control over: “I mean the staff that I’ve got on the counter have been with me for years, and they will come in and say we haven’t seen Brian. If they tell you this bloke hasn’t been in for a week or so, you can do something about it” (Branch Manager D).

The interview data suggest that the more experienced managers found ways of overcoming the constraints of the Balanced Scorecard. They did not allow constraints, such as the lack of detailed and actionable data, to hamper their ability to act. In essence, their experience enabled them to identify ways of overcoming these constraints.

We wanted to explore the idea that more experienced managers were

Table 9
Years of Service

	Sales		Trading Profit	
	(1)	(2)	(3)	(4)
Interaction is evaluated at the mean level of service = 6.6 years:				
Service × (BSC: Aug02–Jul03)	1,724 (909)*	798 (1,014)	401 (203)*	110 (232)
Service × (BSC: Aug03–Jul04)	3,181 (1,229)**	2,458 (1,413)*	659 (220)***	192 (262)
Service × (Aug04–Jul05)	3,857 (1,479)**	3,026 (1,662)*	459 (222)**	48 (270)
Interaction is evaluated at the mean level of service = 11.3 years:				
Senior service × (BSC: Aug02–Jul03)		589 (388)		185 (69)***
Senior service × (BSC: Aug03–Jul04)		460 (529)		297 (92)***
Senior service × (Aug04–Jul05)		528 (618)		261 (117)**
BSC: August 2002 to July 2003	8,320 (2,428)***	8,317 (2,487)***	771 (679)	770 (675)
BSC: August 2003 to July 2004	8,939 (3,965)**	8,932 (3,989)**	998 (835)	994 (818)
August 2004 to July 2005	6,938 (4,770)	6,932 (4,772)	−1,016 (921)	−1,017 (904)
R^2	.16	.17	.13	.13

NOTE.—Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Standard errors (in parentheses) are clustered at the branch level. $N = 11,076$ observations on 156 branches over the period August 1999 to July 2005. Constant, month and branch dummies, demand, minimum temperature, and monthly rainfall are included in all regressions. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

better able to utilize the information in the Balanced Scorecard. The number of interviews was not sufficient to allow us to analyze the response systematically. Instead, we used information from the firm's payroll system. We used information on the years of experience of staff in each branch to look at whether more experienced workers, and in particular more experienced managers, responded better to the Balanced Scorecard.

In table 9, we allow the impact of the Balanced Scorecard to vary with the average years of experience of all staff and of senior staff. Before considering these results, we note that a simple correlation in the data is consistent with the idea that the more experienced managers achieve better performance. The correlation between years of managerial experience and profits is 0.36 (significant at the 1% level).

Turning to the results, in the first and second columns of table 9, we see that experience matters. A branch with staff with the average years

of experience (6.6 years) will have a higher level of sales after the Balanced Scorecard was introduced than a branch with all new staff. In the third and fourth columns, we look at trading profits and show that it is the years of experience of senior staff that matters most. A branch with senior staff with the average years of experience (11.3 years) will have a higher level of trading profits after the Balanced Scorecard was introduced than a branch with all new senior staff.

In results not shown (available from the authors on request), we include the average age of all staff and of senior staff interacted with the Balanced Scorecard indicators and show that it is experience, not age, that is important. Also in results not shown (available from the authors on request), we show that, if we simply split the sample on whether the manager has more or less than 10 years experience (approximately the median), we see all of the effect of the Balanced Scorecard in the greater-than-10-years-experience group.¹⁹

Because this is a key result of this article, we illustrate it further with two figures. In figure 6, we plot the monthly dummies (smoothed over 12 months) from a regression of difference in the level of sales between each Division 1 branch and its matched Division 2 branch, including branch fixed effects, seasonal fixed effects, and weather. The dotted line is the same as appears in figure 1. The top line (long dashes) is for branches where the senior staff have an above-average level of experience. The middle dotted line shows the smoothed dummies when we use all branches. The lower solid dark line shows the smoothed dummies when we estimate using just branches where the senior staff have a below-average level of experience.

The figure shows that, prior to the introduction of the Balanced Scorecard in August 2002 (200208), the difference in sales between Division 1 and Division 2 branches was fairly constant for all groups, with branches with more experienced senior staff having a higher relative level of sales. After the introduction of the Balanced Scorecard, sales in Division 1 branches with more experienced senior staff grew more rapidly, while sales in branches with less experienced senior staff did not. When the Balanced Scorecard was put on hold, the difference in the level of sales evened out in all groups. What is very noticeable from this figure is that the variance in sales performance across branches increased substantially after the introduction of the Balanced Scorecard.

Figure 7 repeats this exercise for trading profits. The figure shows that,

¹⁹ The coefficients (standard errors) on the Balanced Scorecard variables are for the branches with managers with less than 10 years experience: (August 2002 to July 2003) 206 (968); (August 2003 to July 2004) 70 (847); (August 2004 to July 2005) -1,850 (796); for the branches with managers with more than 10 years experience: (August 2002 to July 2003) 1,854 (896); (August 2003 to July 2004) 2,218 (846); (August 2004 to July 2005) -1,339 (709).

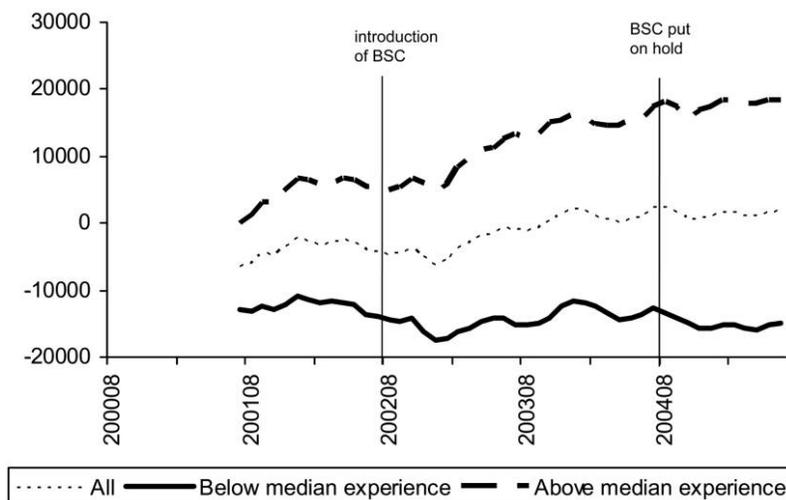


FIG. 6.—Sales in all Division 1 branches relative to Division 2 branches, split by level of experience of senior staff. Each line shows the moving average of monthly time dummies estimated from a regression similar to that in column 2 of table 9, except where the BSC and month dummies have been replaced by 59 individual month dummies that run from August 2008 to the end of the period. The data include 11,076 observations on 156 branches over the period August 1999 to July 2005. The figure shows the 12-month moving average of the 59 estimated coefficients.

prior to the introduction of the Balanced Scorecard in August 2002 (200208), the decline in the relative level of trading profits between Division 1 and Division 2 branches was in branches with less experienced senior staff. After the introduction of the Balanced Scorecard, trading profits in branches with more experienced senior staff grew more rapidly, while trading profits in branches with less experienced senior staff continued to fall. When the Balanced Scorecard was put on hold, the difference in the level of trading profits evened out in all groups.

V. Summary and Conclusions

Our results suggest that the Balanced Scorecard had some impact and that this impact varied significantly by branch. Sales increased on average across all branches, but costs increased by at least as much, so while gross profits did increase, trading profits for the business as a whole did not increase. There is significant variation in the impact. When we separate those branches that perform well on nonfinancial measures from those that perform poorly on nonfinancial performance measures, we find that the first group—those that perform well on nonfinancial measures—experience statistically significant increases in sales, gross profits, and trading profits. One potential explanation of this finding is that branches that

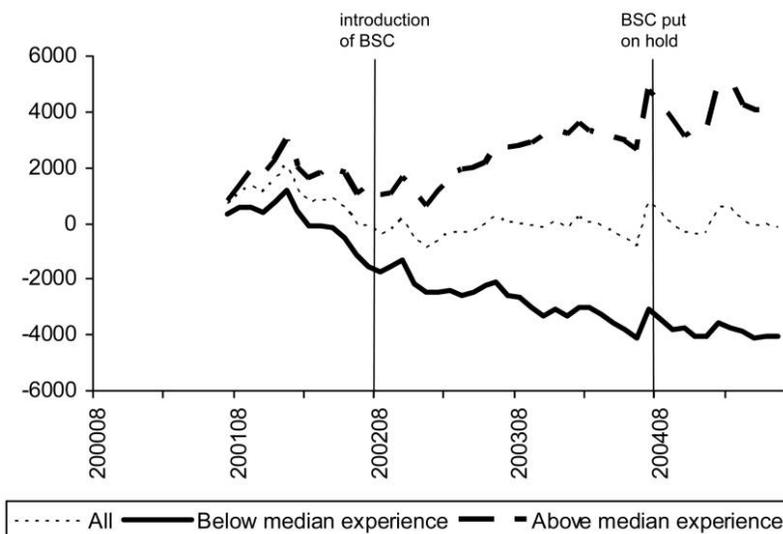


FIG. 7.—Trading profit in all Division 1 branches relative to Division 2 branches, split by level of experience of senior staff. Each line shows the moving average of monthly time dummies estimated from a regression similar to that in column 2 of table 9, except where the BSC and month dummies have been replaced by 59 individual month dummies that run from August 2008 to the end of the period. The data include 11,076 observations on 156 branches over the period August 1999 to July 2005. The figure shows the 12-month moving average of the 59 estimated coefficients.

perform well on nonfinancial measures and financial measures are simply well managed. Note that we are allowing for each branch to have a different average level of performance (we include branch fixed effects). To consider this further, we split the sample into branches that perform relatively well and those that perform relatively poorly on financial measures in the preimplementation period. We then explore whether there are differences in nonfinancial and financial performance in the implementation period. We find that, regardless of prior performance, branches that perform well on nonfinancial measures also perform well on financial measures. This finding is particularly important, as it suggests that the Balanced Scorecard, when implemented correctly and adopted by the branches, has a positive impact on branch performance in terms of sales, gross profit, and net profit.

We then show that years of experience is an important factor in explaining these differences in performance. More experienced managers were able to improve performance. We interpret this as suggesting that the information content of the Balanced Scorecard is what is important, not the incentive per se.

Crucially, the research reported in this article suggests that multi-

dimensional performance measurement systems can have a positive impact on a firm's financial performance providing that the measures and the associated data are presented in a way that enables managers and staff to act on them. It is not enough to introduce an incentive scheme that relates to the measures—the incentive scheme also needs to be accompanied by data that are meaningful to those who have to manage and improve performance. Too often, when performance measurement systems and associated incentive schemes are introduced, they are not designed to enable staff at the front line to take action. The schemes are too complex and contain data that are not sufficiently disaggregated to enable action to be taken. Determining how to design measurement and incentive schemes configured for action is a significant challenge for future research.

References

- Acemoglu, Daron, Philippe Aghion, Claire Lelarge, John Van Reenen, and Fabrizio Zilibotti. 2007. Technology, information and decentralization of the firm. *Quarterly Journal of Economics* 122, no. 4:1759–99.
- Andrews, J. Craig. 1988. Motivation, ability, and opportunity to process information: Conceptual and experimental manipulation issues. In *Advances in consumer research*, ed. Michael J. Houston, vol. 15, 219–25. Provo, UT: Association for Consumer Research.
- Argyris, Chris. 1952. *The impact of budgets on people*. New York: Controllership Foundation.
- Baker, George. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* 100, no. 3:598–614.
- Baker, George P., Michael C. Jensen, and Kevin J. Murphy. 1988. Compensation and incentives: Practice versus theory. *Journal of Finance* 43, no. 3:593–616.
- Banker, Rajiv, Gordon Potter, and Dhinu Srinivsan. 2000. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *Accounting Review* 75, no. 1:65–92.
- Batra, Rajiv, and Michael Ray. 1986. Situational effects of advertising repetition: The moderating influence of motivation, ability, and opportunity to respond. *Journal of Consumer Research* 12, no. 4:432–45.
- Blundell, Richard, and Monica Costa-Dias. 2000. Evaluation methods for non-experimental data. *Fiscal Studies* 21, no. 4:427–68.
- . 2005. Alternative approaches to evaluation in empirical microeconomics. Photocopy, University College London. <http://www.econ.ucl.ac.uk/courses/mect2/rbmcd-13-03-2006.pdf>.
- Bourne, Mike, Mike Kennerley, and Monica Franco. 2005. Managing through measures: A study of the impact on performance. *Journal of Manufacturing Technology Management* 16, no. 4:373–95.

- Bresnahan, Timothy, Erik Brynjolfsson, and Lorin Hitt. 2002. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics* 117, no. 1: 39–76.
- Burgess, Simon, Carol Propper, Marisa Ratto, and Emma Tominey. 2004. Incentives in the public sector: Evidence from a government agency. CMPO Working Paper no. 04/103, Centre for Market and Public Organisation, University of Bristol.
- Cook, Thomas, and Donald Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cools, Kees, and Mirjam van Praag. 2003. The value relevance of disclosing a single-valued corporate target: An explorative empirical analysis. Discussion Paper no. 03-049/3, Tinbergen Institute. <http://www.tinbergen.nl/discussionpapers/03049.pdf>.
- Davis, Stan, and Tom Albright. 2004. An investigation of the effect of Balanced Scorecard implementation on financial performance. *Management Accounting Research* 15 (June): 135–53.
- Fink, Ronald. 2004. New carrots, old yardsticks? *CFO Magazine* (June). <http://www.cfo.com/article.cfm.3014075?f=related>.
- Gates, Stephen. 1999. Aligning strategic performance measures and results. Conference Board Report no. 1261–99-RR, Conference Board, New York.
- Gosling, Jonathan. 2003. A worldly mindset: Satisfying the needs of stakeholders. In *Creating an agenda for European leadership: Conference proceedings*, Bled School of Management, June 10–11. <http://www.leadership-studies.com/research/viewabs.asp?id=42>.
- Hayes, Robert, and William Abernathy. 1980. Managing our way to economic decline. *Harvard Business Review* 58 (July–August): 67–77.
- Heckman, James, Hide Ichimura, and Petra Todd. 1997. Matching as an econometric evaluation estimator. *Review of Economic Studies* 64:605–54.
- Holmstrom, Bengt. 1982. Moral hazard in teams. *Bell Journal of Economics* 13, no. 2:324–40.
- Holmstrom, Bengt, and Paul Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics, and Organisation* 7:24–52.
- Hoque, Zahirul, and Wendy James. 2000. Linking Balanced Scorecard measures to size and market factors: Impact on organizational performance. *Journal of Management Accounting Research* 12, no. 1:1–18.
- Ittner, Christopher D., David F. Larcker, and Marshall W. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a Balanced Scorecard. *Accounting Review* 78, no. 3:725–58.
- Jensen, Michael. 2001. Value maximization, stakeholder theory, and the corporate objective function. *Journal of Applied Corporate Finance* 14, no. 3:8–21.

- Kaplan, Robert S., and David P. Norton. 1992. The Balanced Scorecard: Measures that drive performance. *Harvard Business Review* 70, no. 1: 71–79.
- Kerr, Steven. 1995. On the folly of rewarding A while hoping for B. *Academy of Management Executive* 9, no. 1:7–14.
- Larkin, Ian I. 2006. The cost of high-powered incentives: Salesperson gaming in enterprise software. Working paper, Harvard Business School.
- Lazear, Edward. 2001. Performance pay and productivity. *American Economic Review* 90, no. 5:1346–61.
- Lazear, Edward, and Kathryn Shaw. 2007. Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives* 21, no. 4:91–114.
- Lemieux, Thomas, W. Bentley Macleod, and Daniel Parent. 2007. Performance pay and wage inequality. NBER Working Paper no. 13128 (May), National Bureau of Economic Research, Cambridge, MA.
- MacInnis, Deborah, Christine Moorman, and Bernard Jaworski. 1991. Enhancing and measuring consumers' motivation, opportunity, and ability to process brand information from ads. *Journal of Marketing* 55 (October): 32–53.
- Maisel, Lawrence. 2001. *Performance measurement practices survey results*. New York: American Institute of Certified Public Accountants.
- Malina, Mark, and Frank Selto. 2001. Communicating and controlling strategy: An empirical study of the effectiveness of the Balanced Scorecard. *Journal of Management Accounting Research* 13:47–90.
- Marr, Bernhard, Andrew Neely, Monica Franco, Mark Wilcox, Chris Adams, and Steve Mason. 2004. Business performance measurement: What is the state of the art? In *Proceedings of the fourth International Performance Measurement Association conference*. Edinburgh: International Performance Measurement Association.
- Neely, Andrew, Mike Kennerley, and Veronica Martinez. 2004. Does the Balanced Scorecard work? An empirical investigation. In *Proceedings of the fourth International Performance Measurement Association conference*. Edinburgh: International Performance Measurement Association.
- Neely, Andrew, Bassill Yaghi, and Nigel Youell. 2008. Enterprise performance management: The global state of the art. Working paper, Cranfield School of Management.
- Nørreklit, Hanne. 2000. The balance on the Balanced Scorecard: A critical analysis of some of its assumptions. *Management Accounting Research* 11, no. 1:65–88.
- . 2003. The Balanced Scorecard: What is the score? A rhetorical analysis of the Balanced Scorecard. *Accounting, Organisations, and Society* 28, no. 6:591–619.
- Oyer, Paul. 2004. Why do firms use incentives that have no incentive effects? *Journal of Finance* 59, no. 4:1619–50.

- Parker, Sharon, and Toby Wall. 1998. *Job and work design: Organizing work to promote well-being and effectiveness*. Thousand Oaks, CA: Sage.
- Prendergast, Canice. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (March): 7–63.
- Ridgway, V. F. 1956. Dysfunctional consequences of performance measurements. *Administrative Science Quarterly* 1, no. 2:240–47.
- Rigby, Darrell. 2001. Management tools and techniques: A survey. *California Management Review* 43, no. 2:139–60.
- Speckbacher, Gerhard, Jurgen Bischof, and Thomas Pfeiffer. 2003. A descriptive analysis on the implementation of Balanced Scorecards in German-speaking countries. *Management Accounting Research* 14, no. 4:361–87.