Outcome effect, controllability and performance evaluation of managers: some field evidence from multi-outlet businesses

Dipankar Ghosh*, Robert F. Lusch

University of Oklahoma, Price College of Business, 397W Brooker, Room 200, Norman OK, 73109, USA

Abstract

This field study provides evidence of the outcome effect in performance evaluations of managers in an organization. Specifically, in a retail chain, subjective evaluations of store managers by their supervisors were negatively impacted by unfavorable outcome knowledge. As expected, outcome determinants over which the managers have control influence their performance evaluations and environmental determinants of outcome over which they have no control do not influence their evaluations. However, unexpectedly, central management determinants of outcome over which the managers have no control also influence their evaluations. After these outcome determinants are considered, we find evidence of an outcome effect since failure of the store to meet its target outcome results in a more negative performance evaluation of the managers. Also, the extent to which store managers' evaluations are prone to the outcome effect is not contingent on the measure of the outcome used. © 2000 Elsevier Science Ltd. All rights reserved.

Evaluating managerial performance is an important function of control in organizations. Both subjective (e.g. ratings of subordinates' performance by their superior) and objective (e.g. profitability) criteria are used in these evaluations. However, it is well known that the subjective evaluations of managers are correlated with objective measures of outcomes. For example, a meta-analysis by Bommer, Johnson, Rich, Podsakoff and Mackenzie (1995) of studies containing both objective and subjective ratings of employee performance resulted in a mean correlation of 0.39. Clearly the manager should be given some credit for the objective performance of the organization that a manager helps to achieve. However, when subjective measures of managerial performance are influenced by objective measures of performance that the manager does not control, then the performance evaluation system is less than ideal and therefore the control system is suboptimal.

One of the main reasons for this weakness in performance evaluation and/or control systems is the "outcome effect" or bias in which evaluators assess manager's performance based on outcomes (Baron & Hershey, 1988; Hawkins & Hastie, 1990). Thus, when an outcome is positive (negative), evaluators tend to evaluate the manager positively (negatively) regardless of the actual appropriateness of the decision resulting in that outcome. By taking outcomes into account that do not reflect managerial performance will affect not only the quality of the evaluation decision but will also incorrectly reward/penalize managers.

As much of the research studying the outcome effect has been undertaken in a laboratory setting,
the primary objective of this research is to assess if that effect can be shown to exist in an actual organization. In addressing this problem it was necessary to develop a methodology for studying the outcome effect in the field. Importantly, the development of this methodology was a secondary objective of this research. The development of this methodology is more complex than developing a field study questionnaire. It was necessary to gather data from four distinct sources and the data had to be integrated into a two-step mathematical model that was used to isolate the presence or absence of the outcome effect. The methodology is applicable to a multi-unit business and, thus, can be used as the basis for future studies interested in addressing how a host of organizational factors could alleviate the outcome effect. In general, the methodology can be used to obtain a better understanding of the outcome effect and whether its influence can be mitigated during the manager's performance evaluation.

The research specifically addresses two questions: (i) Are subjective performance evaluations of segment managers influenced by outcomes after controllability of these outcomes are taken into consideration? (ii) Are outcome effects robust across alternative measures of segment economic outcome, namely, sales and gross profit?

The research questions are motivated by several reasons. First, the outcome effect has been documented by a large number of laboratory studies (Baron & Hershey, 1988; Brown & Solomon, 1987, 1993; Lipe, 1993; Lipshitz, 1989; Mitchell & Kalb, 1981) which with their high internal validity provides the theory the best chance of being supported. But when a research design emphasizes one type of validity, it necessarily weakens the other (Abdel-khalik & Ajinkya, 1979). Since the two primary criteria for evaluating the quality of research are internal and external validity, the next step, therefore, is to test the theory in a field setting with its high external validity. The triangulation of research, testing the outcome effect in both the laboratory and a field setting, can demonstrate the robustness of the theory.

Second, evaluating managers in organizations is more complicated than the outcome effect literature implies since the laboratory context is substantially different from the situation faced by an evaluator in an organization (Borman, 1992). For example, in experiments the decision maker is usually evaluated on a single decision where the connection between the decision and the outcome is rather transparent and the duration between the two is relatively short. In contrast, managers in organizations are evaluated for their overall performance which involves a host of interrelated decision variables collectively impacting on outcomes over varying time periods. Furthermore, since the evaluator is often not completely aware about the manager's decision environment, how the decisions impact the outcomes is not transparent. For example, outcomes are commonly delayed and not easily attributable to a particular action and variability in the environment degrades the reliability of outcome feedback (Tversky & Kahneman, 1986). Also, evaluative decisions are rarely context-free. Rather, a broader sample of evidence with various degrees of objectivity and consistency is available, some of which may not pertain to the manager's performance because the manager has no control over them.

Finally, prior research assessing the outcome effect did not control for “uncontrolled” factors. Decision researchers (and most other people) would argue that good decisions tend to lead to good outcomes and bad decisions lead to bad outcomes. Thus, in the absence of other information about the initial decision, using outcomes to evaluate the decision is reasonable and, therefore, an outcome effect is not established by merely showing that evaluators take outcomes into account (Hershey & Baron, 1992). However, this is not the case when controllability of outcomes is considered. When an outcome is controllable by the manager then the manager should be held responsible for the consequences of his or her decisions (Kelley & Michela, 1980; Tan & Lipe, 1997). Similarly, a manager should not be held responsible for the consequences of his or her decisions when an outcome is not controllable by them.

This field study, conducted in a retail chain organization, provides evidence of an outcome effect in performance evaluations. Outcome was measured in two ways: (i) a continuous economic measure (sales and gross profit), and (ii) a dichotomous economic
measure (knowledge of whether the target sales and gross profit for the period was met or not met). Furthermore, a methodology was developed to partition the continuous economic measures into outcome under the control of store managers and outcome not under their control. If we look at our continuous measure of outcome (sales and gross profit), as expected, the results show that sales and gross profit over which managers have control influence how they are evaluated by their superiors. Also as expected the sales and gross profit outcomes determined by the environment over which the managers have no control do not influence their evaluations. However, unexpectedly the results also show that sales and gross profit outcomes controlled by the central management and over which the managers have no direct control also influences manager’s evaluations. This is an unexpected effect; however, we provide a plausible alternative explanation in the discussion section. When we study the influence of the dichotomous measure of outcome (i.e. knowledge of target sales and gross profit achieved) we find that subjective evaluations of store managers were negatively impacted. Given that we have statistically accounted for the outcomes managers have control or no control over this finding provides strong evidence of an outcome effect. Finally, this study found that the outcome effect is generally not contingent on the two different economic measures of segment outcome.

The remainder of the paper is organized as follows. The next section reviews the literature and develops the hypotheses. Section three discusses the research method, section four presents the data analyses, and section five reviews the results and examines the implications of this study.

1. Development of research hypotheses

1.1. Performance evaluation and outcome effect

As stated above, the outcome effect has been empirically demonstrated in several experimental studies. Mitchell and Kalb (1981) investigated the outcome effect in supervisor’s evaluations of subordinates in a health care setting. The study found that those supervisors with outcome knowledge, especially in the case of a negative outcome, rated the outcome as more probable, held the subordinate more responsible for the outcome, and made more internal attributions for the outcome than subjects did with no outcome knowledge. Baron and Hersey (1988), in a series of five studies, gave subjects a set of 12–16 medical and gambling decisions to evaluate and the outcome of each of the decisions. The results were very consistent: the outcome of the decision (good or bad) systematically influenced subjects’ evaluations of the quality of the decision.

Accounting research has tried to isolate factors that can eliminate the outcome effect by making the evaluator familiar with the manager’s decision environment. This familiarity can facilitate evidence recall (other than the outcome) during the manager’s performance evaluation (Fischhoff, 1975). For example, Brown and Solomon (1987, 1993) tried to enhance the involvement of the evaluator in the manager’s (i.e., the evaluatee’s) decision by acquainting the evaluator with the manager’s decision variables. In their first study, evaluators were asked some questions about the manager’s task prior to making the performance evaluation. In a subsequent study, the evaluators were made “advisors” to the managers and had to have a decision consensus between themselves and the managers. In Fisher and Selling (1993), the evaluator observed the decision making process of the managers. Finally, in Lipe (1993), expenditures on variance investigation were framed for the evaluator as either a cost or a loss, depending on whether the investigation did or did not find a problem.

In organizations, both subjective and objective measures are used to assess managers’ performance. Subjective measures often involve multi-item instruments where supervisors, peers, or others familiar with the incumbent’s work assess his or her performance. It has been frequently demonstrated that these subjective measures are prone to biases like the outcome effect. However, to presume that objective criteria are better can be misleading. Objective measures are notoriously deficient as criteria because they typically tap only a small proportion of a manager’s job performance requirements (e.g. Guion, 1965). Further, contamination can also be a problem with some of the
objective criteria as well (Borman, 1992). Thus, subjective measures, in spite of their limitations, are most frequently used in manager’s performance evaluation (Borman, 1992).

The reason why we observe the outcome effect in manager’s evaluations is because outcome knowledge influences the evidence recalled by the evaluator when attempting to assess the performance of the manager (Slovic & Fischhoff, 1977). Evidence recall is a deliberate task and since humans have limited information processing capacity (Simon, 1957), “knowing what happened facilitates knowing where to look for and what to accept as reasons” (Fischhoff, 1975, p. 297). Thus, as observed in the experimental studies, the outcome effect should persist in organizations although, as we discuss next, using outcomes to evaluate performance is not always reasonable.

1.2. Outcome effect and outcome controllability

In organizations, managers have varying levels of control over outcomes. In which case it may be reasonable to use outcome knowledge for evaluations since they provide us with information about manager’s decisions (Baron & Hershey, 1988). But to make the outcome more informative, one must separately model how much the outcomes of the segment are due to actions and decisions of segment managers. Hence, we next identify the decision factors affecting a segment’s outcomes and examine whether those factors, grouped by their controllability, influence the segment manager’s subjective performance evaluation. If after accounting for these factors the manager’s subjective performance evaluation continues to be affected by knowledge of whether the target outcomes were achieved, we will have evidence of an outcome effect.

When is it reasonable to use outcome knowledge? The distinction between a good decision and a good outcome is a basic one in all decision analysis because they are not the same (Edwards, 1984). Some argue that using outcome knowledge may be justified and appropriate in some contexts while remaining erroneous in others. In general, when the manager has, or ought to have, more information than the evaluator, outcome is diagnostic of the quality of the manager’s decisions since it is reasonable to assume that bad (good) outcomes results from poor (good) decisions (Baron & Hershey, 1988). However, if there is no information asymmetry problem, an evaluator should only use the information about potential outcomes and the probabilities and utilities of those outcomes that existed at the time of the decision in evaluating decision quality (Hershey & Baron, 1992).1 Hershey and Baron (1995) specifically addressed the issue of when is it reasonable (or “normative,” using their term) to use outcome knowledge and illustrate that this knowledge is sometimes informative about decision quality. For example, using their terminology, assume the evaluator is trying to determine if the manager made a good decision (i.e. G) or a bad decision (i.e. B) and the only knowledge available to the evaluator is the outcome resulting from the decision, namely, success (i.e. S) or failure (i.e. F). The outcome knowledge can then be used to infer the manager’s decision quality provided the following relations among the conditional probabilities hold.

$$P(S | G) > P(S | B) \text{ and } P(F | B) > P(F | G)$$

Hershey and Baron (1995) concludes “…that the likelihood of a good decision is increased by the occurrence of a success, and the likelihood of a bad decision is increased by the occurrence of a failure” (p. 127). Thus, outcome knowledge can be very informative about the decision quality if that decision significantly influences outcome probability.

Further, the extent to which outcome knowledge is informative about the decision is directly related to the degree by which its probabilities are increased by the decision. Thus, when $P(S|G)$ is considerably higher than $P(S|B)$, and $P(F|B)$ is considerably higher than $P(F|G)$, outcome knowledge is more informative regarding decision quality. Such would be the case if the manager has high control over the factors influencing outcome since controllability.

---

1 This is consistent with the agency theory perspective (Holmstrom, 1979) that when the agent’s effort is observable, the first best solution is to reward the agent based upon those efforts. However, if the agent’s effort is either unobservable or imperfectly observable, the best solution is to reward the agent upon outcomes alone or some combination of outcomes and information that is incrementally revealing of effort.
increases the responsibility attached to a manager for the consequences of his or her decisions (Kelley & Michela, 1980). Specifically, when a manager has high control over the outcome, then \( P(S|G) \) will be high with \( P(S|B) \) low due to the perceived direct and causal (i.e. controllable) relation between the decision and the outcome. Likewise, \( P(F|B) \) will be high with \( P(F|G) \) low. Thus, in these situations, the outcome knowledge is more informative regarding the quality of the manager’s initial decision. On the other hand, with a more uncontrollable outcome, these direct causal relations are absent and it is difficult to assess the relative size of the relevant probabilities. In that case, outcome knowledge is much less informative regarding the manager’s initial decision. Thus, outcome controllability can be a criterion not only to establish the reasonableness of the use of outcome knowledge in evaluating decision quality but also indicative of when use of outcome knowledge may be observed in evaluations.

Before proceeding further, we need to better understand the term controllability. Controllable decision factors are those over which the segment manager has considerable latitude. For example, often segment managers can control such things as marketing, production schedules, and pricing. Obviously the actual research setting will determine which factors are controllable. Uncontrollable factors are the environmental factors and the central office factors affecting a segment’s outcome which a manager cannot control. These factors can include both supply and demand factors that are in the firm’s external environment and resource support from central management such as facilities and equipment.

Therefore, the above discussion suggests that when outcome factors are controllable by the segment manager, it should be reasonable for the evaluator to consider these factors in evaluating the segment manager’s performance. This is consistent with Hershey and Baron’s (1995) mathematical results that the conditional probability that is expected under conditions of controllability, leads correctly to use of the outcome knowledge. On the other hand, with regard to factors uncontrollable by the segment manager, considering these factors for evaluation should not be considered as reasonable. Thus, once the influence of the outcome factors grouped by their controllability are accounted for, theoretically, the outcome knowledge should be inconsequential to manager’s subjective performance evaluation. However, as discussed earlier, experiments have consistently demonstrated that subjective performance evaluations are biased by the outcome knowledge. For example, even when the evaluators had all the relevant information known to the evaluatee, plus the outcome knowledge, the evaluators took outcome into consideration in rating the quality of the decision of the evaluatee (Baron & Hershey, 1988). The preceding discussion allows us to state our hypotheses.

**H1A**: Segment manager’s subjective performance evaluation will be affected by sales and gross profit outcomes they control and will not be affected by sales and gross profit outcomes they do not control.

**H1B**: After accounting for the controllability of sales and gross profit outcomes, the segment manager’s subjective performance evaluation will be affected by knowledge of whether the target sales and gross profit goal was achieved (i.e. target outcome knowledge).

With this discussion of theory, hypotheses, and definitions we now move to a more concrete discussion of the specific research setting and method for hypotheses testing.

### 2. Research method

#### 2.1. Overview of the research setting

The research setting selected is a multi-unit retail chain store organization where each store can be viewed as a separate business unit with a separate manager (referred to as the store manager which is analogous to the “segment manager” we refer to in our hypotheses). The store manager’s performance evaluation is a subjective assessment conducted by the regional manager. As we will see below in a retail store setting, we can isolate the determinants of store sales and gross profit. Since some of these determinants are controllable and others uncontrollable by the store manager, we
can isolate the outcomes (sales and gross profit) controllable by the store manager. There are two reasons why the retail chain in this study provides an appropriate field setting to detect if outcome effects occur in the more complex setting of an actual organization. First, each store is a decentralized segment of a retail chain as well as a distinct and stand alone economic unit. Thus, store-specific accounting information is less noisy and factors affecting each store’s outcome are more easily identified. Second, the regional managers who evaluate the performance of the store managers were once store managers themselves. Further, they very closely monitor each store under their supervision, including frequent store visits. Thus, it can be assumed that by virtue of their experience, the regional managers were familiar with the factors that significantly influence store outcomes (Davis, 1982; Choo, 1989).

2.2. Data sources and data collection

The data were collected from a chain store retailer with almost 250 stores in the United States. The statistical analysis was based on 204 stores for which complete data were available for all variables. The retail units are located in five major metropolitan statistical areas (MSA) and rural towns in nearly equal numbers. Most stores have 3000–10,000 square feet of selling area while more recent units are larger, ranging from 15,000 to 25,000 square feet of selling space. Each store offers a wide range of general merchandise items, catering to a relatively narrow trading area with a tight customer focus. Most stores are fairly uniform in terms of their merchandise mix and store layouts. The stores are staffed by a store manager as well as one or more assistant managers. A regional manager monitors the operations of 10–30 retail units by regularly visiting the stores and reviewing reports on the financial performance of the units. In addition, each regional manager has previously served as a store manager in this organization and, thus, is familiar with the decisions and control the manager has over operations.

The data were collected from four sources. First, the relevant economic and operating characteristics of each store were obtained from the internal accounting records. For each store, there were income statements, selected balance sheet data and information about other resources. For this study, the following store-specific data items were obtained: annual sales, gross profit, average inventory investment, annual advertising expenditures, gross profit percentage, number of full time employees, total square feet of space for the store and total square feet of selling space for the store, and if the store’s target annual sales and gross profit were met. Second, we were able to obtain the internal records data on the regional manager’s annual rating of the store manager’s effectiveness or performance. Third, supply related environmental factors, namely, the size of the market, the size in square feet of the store’s number one competitor, the number of competing stores in the trade area or store saturation, and the economic health of the trade area, were obtained. These data were obtained from a measurement instrument designed jointly by the researchers and the retail organization to provide information on the nature and scope of competition in each trade area as well as the overall economic health of the trade area in which the store was located. A form was sent to each store manager requesting information on competitors. This form was provided through the normal organizational channels to each manager with a cover letter and a set of instructions. Each returned form was reviewed by the researchers and the regional manager for missing or miscoded items. If required, the form was returned to the store manager to ensure complete and accurate responses. Finally, the demand related environmental factor, namely, the market match which depended on the population mix of the trade area, was purchased from National Planning Data Corporation. This firm commercially supplies demographics, based on U.S. Census of Population and Housing, for trade areas, as defined by the client.

2 The instrument contains proprietary information. Further, the organization expended considerable resources to develop it. Hence, the organization was unwilling to make the complete instrument publicly available. However, a detailed description of the questions from the instrument relevant for this research is included in the discussion of the “Independent Variables” of Eq. (2).
2.3. Segment (i.e. store) outcomes

Store outcome was gauged using two measures: annual net sales per square foot of selling space (hereafter, referred to as SALES) and annual gross profit per square foot of selling space (hereafter, referred to as GPROFIT). A review of measures of store performance revealed that they are both widely used measures of store performance (Lusch, 1986) and are frequently reported in retail trade publications and operating result studies by retail trade associations (National Retail Hardware Association, 1992). Each store also has target outcomes for annual sales and gross profit. For analysis purpose, knowledge of target outcome achievement was coded as ‘0’ for target success and ‘1’ for target failure.

2.4. Determinants of segment outcomes

2.4.1. Environmental factors

We identified four uncontrollable environmental factors which influence segment outcomes. These factors represent both supply and demand related factors. Market matching is the extent to which the demographics of the trade area match the desired target market. Thus, one would expect that a higher degree of market matching leads to higher economic performance of a store. The market matching concept is very similar to the concept of environment-strategy coalignment in the strategic management literature (Venkatraman & Prescott, 1990). In this study, market matching was indicated by the percentage of Hispanics ($E_1$) in the trade area population since Hispanics were the primary target market of the stores. Relative store size of No. 1 competitor ($E_2$) was measured by dividing the square footage of the largest trade area competitor by the square footage of our sample store. Huff’s (1964) retail patronage and trade area model prominently recognizes the importance of the relative size of a retail store or center in attracting patrons. The higher the relative store size of the No. 1 competitor, then the lower the economic performance of the store. Store saturation ($E_3$) is another major environmental variable influencing store performance and it is measured by the number of stores in the retailer’s trade area per thousand population in the trade area (Ingene & Lusch, 1981). The economic performance of the store is negatively correlated with store saturation. Finally the economic health of the trade area ($E_4$) is measured on a seven-point scale, measuring whether it is a good time to open a store in the trade area. Responses could range from “definitely not”(1) to “definitely yes”(7). If it is believed that opening an additional store would be a good decision, then the economic health of the trade area is likely to be good. One would expect that the better the economic health of the trade area, the higher the economic performance of the store.

2.4.2. Central management factors

For the retail chain organization under study, the first central management factor is store size ($C_1$), measured by the total square footage of the store. The second factor, size of market ($C_2$), is indicated by whether the store was located in a metropolitan statistical area (MSA) or nonmetropolitan environment (i.e. rural town). If a store was located in a nonmetropolitan market, then it was coded a ‘1’ and ‘0’ otherwise. Size of market was important because the chain stores in this retail organization tended to operate better in nonmetropolitan environments where fewer competitors are located. Both store size as well as the location is decided by the central management and hence are uncontrollable by the store manager.

2.4.3. Store manager factors

The store manager has control on four important decision factors — inventory, advertising, customer service capacity and pricing. Inventory intensity ($D_1$) is measured by inventory investment per square foot of selling area. This is a frequently mentioned determinant of store performance (Lusch, Dunne & Gebhardt, 1993). The higher the inventory intensity, the higher the economic performance of the store. Advertising intensity ($D_2$) is measured as annual advertising expenditures per square foot of selling area. The higher the advertising intensity, the higher the economic performance of the store. Customer service capacity ($D_3$) is measured as square feet of selling space per full time equivalent employee. This is a popular measure of
customer service capacity (Ingene & Lusch, 1981). However, it should be noted that a high value on this measure (i.e. a high number of square feet of selling space per employee) is an indicator of low service capacity while a low value (i.e. relatively few square feet of selling space per employee) is a measure of high service capacity. The impact of customer service capacity on the economic performance of the store probably depends on the type of store and line of retail trade. For the retail chain organization under study, stores are positioned as neither self service nor full service but somewhere between these extremes. Thus, some level of customer service capacity is important. Consequently, one would expect the economic performance of the store to be positively influenced by customer service capacity. Finally, store performance is affected by the gross margin percent ($D_4$) which Lusch and Moon (1984) point out to be a critical decision variable in retailing. Since all store managers are able to order merchandise as they see fit from the chain’s distribution center, they can determine the appropriate mix of merchandise and the initial storewide markup. The store managers, by ordering the “right” quantity and mix of merchandise, also influence the need for markdowns and determine the timing of markdowns. As a consequence, store managers have a significant influence on the realized gross margins for the store. The effect of gross margin on store economic performance could be either positive or negative, depending on the price elasticity of demand.

2.5. Subjective performance evaluation of segment manager

The measure of store manager performance was operationalized by summing the subjective ratings \([R]\) on eight questions intended to appraise managerial performance (both decision quality and effort). The eight items reflect the supervisor’s satisfaction with the manager’s job performance on the following attributes: managerial skills, taking responsibility, level of motivation, decision making ability, knowledge of trade area, tolerance for pressure, relations with supervisors, and managerial potential. The district supervisor rated each manager on each attribute on a seven-point Likert type scale (where 1 = “very dissatisfied” and 7 = “very satisfied”). We summed the eight ratings to obtain an index with higher scores representing a higher evaluation of the store manager’s overall performance by the supervisor (the scores could range from 8 to 56). Factor analysis of the eight items provides support for this approach. It revealed that they all highly loaded on one factor with an eigenvalue of 5.74 and which explained 71.8% of the standardized co-variances among the eight items.

3. Statistical analysis

Table 1 shows the descriptive statistics and the correlation matrix of all the research variables.

3.1. Hypotheses testing of H1A and H1B

To provide a statistical test for the two hypotheses requires several steps. To begin with, we model the outcome determinants of both SALES and GPROFIT. We propose the following regression model for operationalizing the determinants for both SALES and GPROFIT, the two measures of a store’s economic outcome:

\[
\text{S or } \text{GP} = A_1 + A_2(E_1) + A_3(E_2) + A_4(E_3) \\
+ A_5(E_4) + A_6(C_1) + A_7(C_2) + A_8(D_1) \\
+ A_9(D_2) + A_{10}(D_3) + A_{11}(D_4) + e
\]

(1)

where:

- \(S\) = SALES
- \(GP\) = GPROFIT
- \(E_1\) = market matching No.
- \(E_2\) = relative size of 1 competitor
- \(E_3\) = store saturation
- \(E_4\) = economic health of trade area
- \(C_1\) = store size in square feet
- \(C_2\) = size of market
- \(D_1\) = inventory intensity
- \(D_2\) = advertising intensity
- \(D_3\) = customer service level
- \(D_4\) = gross margin percent
Table 1
Descriptive statistics of the research variables

<table>
<thead>
<tr>
<th>Panel A: Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

Panel B: Correlation coefficients (p-values) [Note: Pearson, except Spearman for (C₃) (Tₛ) and (T₉GP)]

| (S) | (GP) | (E₁) | (E₂) | (E₃) | (E₄) | (C₁) | (C₂) | (D₁) | (D₂) | (D₃) | (D₄) | (R) | Tₛ | T₉GP |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| (S) | 1.00 | 0.98 | 0.29 | 0.07 | −0.05 | −0.36 | −0.11 | −0.13 | 0.48 | 0.47 | −0.69 | 0.25 | 0.33 | −0.24 | −0.18 |
| (GP) | 1.00 | 0.30 | −0.06 | −0.04 | −0.36 | −0.14 | 0.15 | 0.43 | 0.42 | −0.69 | 0.43 | 0.37 | −0.28 | −0.24 |
| (E₁) | 1.00 | −0.03 | −0.22 | −0.03 | 0.16 | −0.15 | 0.15 | −0.12 | −0.20 | 0.16 | 0.19 | −0.11 | −0.11 |
| (E₂) | 1.00 | −0.02 | 0.13 | −0.26 | −0.25 | 0.09 | 0.15 | −0.11 | −0.05 | 0.15 | 0.04 | 0.06 | 0.06 |
| (E₃) | 1.00 | −0.15 | 0.03 | −0.20 | 0.03 | −0.06 | 0.13 | 0.04 | 0.05 | −0.04 | −0.02 |
| (E₄) | 1.00 | −0.18 | 0.13 | 0.22 | 0.11 | −0.35 | 0.13 | 0.11 | −0.17 | −0.16 |
| (C₁) | 1.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.11 | 0.06 | 0.13 | 0.05 | 0.00 |
| (C₂) | 1.00 | 0.15 | 0.25 | 0.07 | 0.08 | 0.03 | 0.01 | 0.00 | 0.96 | 0.84 |
| (D₁) | 1.00 | 0.35 | 0.53 | 0.07 | −0.07 | 0.05 | 0.02 | 0.02 | 0.01 | 0.01 |
| (D₂) | 1.00 | 0.37 | 0.05 | 0.05 | 0.05 | 0.03 | 0.06 | 0.02 | 0.02 | 0.02 |
| (D₃) | 1.00 | −0.15 | 0.11 | 0.16 | 0.14 | 0.16 | 0.08 | 0.08 | 0.08 | 0.08 |
| (D₄) | 1.00 | 0.31 | 0.24 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| (R) | 1.00 | −0.56 | −0.52 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| (Tₛ) | 1.00 | 0.69 | 0.00 |
| (T₉GP) | 1.00 |

| a | Descriptive statistics of the sample 204 stores. Correlations in bold are significant at p ≤ 0.05. Variable definitions: S = sales per square foot of selling space; GP = gross profit per square foot of selling space; E₁ = market matching; E₂ = relative size of No. 1 competitor; E₃ = store saturation; E₄ = economic health of trade area; C₁ = store size in square feet; C₂ = size of market; D₁ = inventory intensity; D₂ = advertising intensity; D₃ = customer service level; D₄ = gross margin percent; R = supervisor’s evaluation of the store manager’s performance; Tₛ = Target sales achieved; T₉GP = Target gross profit achieved. |

---

e = error term which captures unmeasured or omitted variables and random effects or uncertainty

A₁ − A₁₁ = regression parameters to be estimated.

e In Eq. (1), the E variables are uncontrollable environmental variables, the C variables are uncontrollable central management factors, and the D variables are controllable store manager’s decision variables. The error term includes the effect of forcing a linear relation between the variables as well as the effect of all omitted and unmeasured variables and random effects.

Regression equations were estimated separately for SALES and GPROFIT. Multicollinearity was not considered a significant problem since the variance inflation factors (VIFs) for all variables were low (i.e. they ranged from 1.15 to 2.28 for
both SALES and for GPROFIT). Table 2 presents the results separately for SALES and GPROFIT; columns 3–5 pertain to SALES and 6–8 pertain to GPROFIT. Overall, models for both SALES ($F = 40.63, p = 0.0001, R^2 = 0.66$) and GPROFIT ($F = 46.16, p = 0.0001, R^2 = 0.69$) are statistically significant. An examination of the predictor variables for SALES indicates that, with the exception of store saturation ($E_3$), all of the variables are statistically significant ($p \leq 0.05$) with the expected signs on the coefficients. For GPROFIT, store saturation ($E_3$) also is not significant at the customary 0.05 level.

In the above step we identified variables which significantly affected SALES and GPROFIT. Next, we separately examine whether these variables

\begin{equation}
S/GP = A_1 + A_2(E_1) + A_3(E_2) + A_4(E_3) + A_5(E_4) + A_6(C_1) + A_7(C_2) + A_8(D_1) + A_9(D_2) + A_{10}(D_3) + A_{11}(D_4) + e
\end{equation}

Table 2
Regression results for factors affecting store outcome

| Independent variable | Predicted sign | S = dependant variable | | | |GP = dependant variable | | |
| | | Coefficient | $T$-stat. | $P$-value$^b$ | | | Coefficient | $T$-stat. | $P$-value$^b$ |
| Intercept | 0 | -6.723 | -0.43 | 0.6771 | | | -22.248 | -3.61 | 0.0004 |
| $E_1$ | + | 0.131 | 3.43 | 0.0004 | | | 0.053 | 3.59 | 0.0002 |
| $E_2$ | - | -0.839 | -2.56 | 0.0057 | | | -0.322 | -2.52 | 0.0063 |
| $E_3$ | - | 0.362 | 1.75 | 0.0810 | | | 0.142 | 1.77 | 0.0790 |
| $E_4$ | + | 1.419 | 3.10 | 0.0011 | | | 0.548 | 3.08 | 0.0012 |
| $C_1$ | + | 0.001 | 3.36 | 0.0005 | | | 0.001 | 3.20 | 0.0008 |
| $C_2$ | + | 5.480 | 2.36 | 0.0097 | | | 1.999 | 2.21 | 0.0142 |
| $D_1$ | + | 1.614 | 4.02 | 0.0001 | | | 0.619 | 3.96 | 0.0001 |
| $D_2$ | + | 13.650 | 4.75 | 0.0001 | | | 4.701 | 4.20 | 0.0001 |
| $D_3$ | - | -0.032 | -7.54 | 0.0001 | | | -0.012 | -7.34 | 0.0001 |
| $D_4$ | + | 0.014 | 4.46 | 0.0001 | | | 0.010 | 8.79 | 0.0001 |
| Adjusted $R^2$ | | 0.661 | | | | | 0.690 |
| Model $F$-value | | 40.627 | | | | | 46.155 |
| Model $P$-value | | 0.0001 | | | | | 0.0001 |

$^a$ S = annual sales per square foot of selling space or SALES; GP = annual profit per square foot of selling space or GPROFIT; $E_1$ = market matching; $E_2$ = relative size of No. 1 competitor; $E_3$ = store saturation; $E_4$ = economic health of trade area; $C_1$ = store size in square feet; $C_2$ = size of market; $D_1$ = inventory intensity; $D_2$ = advertising intensity; $D_3$ = customer service level; $D_4$ = gross margin percentage.

$^b$ All tests are one-tail tests except for gross margin percent ($D_4$) and store saturation ($E_3$).

$^c$ Customer service is measured by square feet of selling area per fulltime equivalent employee. Therefore a high value represents low customer service and a low value represents high customer service levels.

---

3 The variance inflation factor is the reciprocal of $1 - R^2$ when a particular independent variable is regressed on all other independent variables. See Belsley, Kuh and Welsch (1980) for a discussion of this diagnostic approach.

4 The coefficient of the store saturation (or $E_3$) variable, although not statistically significant, is unexpectedly positive. We expected that as the market becomes more saturated, sales per square foot would decline. The president of the retail chain was, therefore, consulted about this finding. His explanation was that in highly saturated markets the stores tended to be viewed as more convenient and accessible since they were relatively small and thus had a differential competitive advantage in these saturated markets. Another explanation, not offered by the president, may be that in highly saturated trade areas a household may be more likely to shop inside the trade area (versus, outside the trade area) and, thus, total market sales rise. At the same time, because of the large amount of retail space in the trade area, households from outside the trade area may visit the trade area to shop for a higher percentage of their purchase requirements. In brief, over-saturated trade areas may actually be attractors of more demand, which increases their market potential.
which affected SALES and GPROFIT also affect the subjective performance evaluation of store managers. To do this we must estimate how much in SALES and GPROFIT is due to the four broad categories of determinants of store outcomes which, as discussed previously, were: (1) environmental factors, (2) central management factors, (3) store manager factors, and (4) all other unmeasured or omitted variables and random effects. If one takes the information incorporated in the regression equations presented in Table 2, one can decompose the sources of SALES or GPROFIT outputs into four components as follows: \( PE \) = predicted outcomes due to uncontrollable environmental factors. This value is obtained by considering only the significant factors from separate regressions for SALES and GPROFIT in Table 2. Thus, for example, from Table 2 for SALES: 

\[
PE = 0.131(E_1 \text{ or market matching}) - 0.839 \text{ (E2 or relative size of No. 1 competitor)} + 1.419 \text{ (E4 or economic health of trade area)}.
\]

For GPROFIT, since the same environmental variables were also significant, we have the following from Table 2:

\[
PE = 0.053(E_1 \text{ or market matching}) - 0.322(E_2 \text{ or relative size of No. 1 competitor}) + 0.548(E_4 \text{ or economic health of trade area}).
\]

\( PC \) = the predicted outcomes due to uncontrollable central management factors. For SALES, this value is obtained from regression for SALES in Table 2:

\[
PC = 0.001(C_1 \text{ or store size}) + 5.480(C_2 \text{ or size of market}).
\]

A similar procedure was also conducted for GPROFIT to obtain:

\[
PC = 0.001(C_1 \text{ or store size}) + 1.999(C_2 \text{ or size of market}).
\]

\( PD \) = the predicted outcomes due to decisions the store manager controls. For SALES, this value is obtained from regression equation in Table 2:

\[
PD = 1.614(D_1 \text{ or inventory intensity}) + 13.650(D_2 \text{ or advertising intensity}) - 0.032(D_3 \text{ or customer service level}) + 0.014(D_4 \text{ or gross margin percentage}).
\]

A similar procedure was also conducted for GPROFIT:

\[
PD = 0.619(D_1 \text{ or inventory intensity}) + 4.701(D_2 \text{ or advertising intensity}) - 0.012(D_3 \text{ or customer service level}) + 0.010(D_4 \text{ or gross margin percentage}).
\]

\( PU \) = outcomes due to all omitted and unmeasured determinants including random effects; this is estimated by \( e \), the residual term from each of the equations in Table 2. This residual term (\( e \)) is the predicted store outcomes less the actual store outcomes based on the regression equations presented in Table 2.

We can now proceed to suggest a test of H1A and H1B. Consider the following regression equation:

\[
R = B_0 + B_1(PE) + B_2(PC) + B_3(PD) + B_4(PU) + B_5T + e
\]

where:

\[
R = \text{the supervisor’s subjective evaluation of the store manager’s performance};
\]

\( PE, PC, PD, PU \) = the decomposed outcome (SALES or GPROFIT) due to different sources as discussed above;

\( T \) = achievement of target outcome for either SALES or GPROFIT (coded as dummy variable; “1” for failure to meet the target and “0” otherwise);

\( e \) = error term;

\( B_1 - B_5 \) = regression parameters to be estimated.

In Table 3 the statistical results for this regression equation are provided separately for both SALES and GPROFIT. Multicollinearity is not considered a significant problem (the VIFs ranged from 1.03 to 1.20 for SALES and 1.01 to 1.31 for GPROFIT). Columns 2–4 present the results for the SALES equation and columns 5–7 present the results for the GPROFIT equation.

As hypothesized in H1A, SALES due to uncontrollable environmental factors (PE) has no influence on the performance rating the manager receives. Similar results are obtained when GPROFIT is the output measure. Also, as hypothesized in H1A, controllable factors affect SALES and GPROFIT (i.e. PD) significantly influences the store manager’s subjective performance evaluation. For example, when SALES (GPROFIT) which is due to these controllable
Decision factors rise or fall by one dollar the manager’s performance rating rises or falls by 0.06 (0.20) points.

Hypothesis 1A also predicted that the manager’s performance will not be affected by central management factors since they are uncontrollable at the store manager’s level. In Table 3 we do not find support for this prediction for either SALES and GPROFIT output. For example, when SALES attributable to central management factors ($PC$) rises or falls by one dollar the store manager gains or loses 0.41 rating points, as the case may be, which is significant at $p = 0.0003$. Similarly, when GPROFIT attributable to central management factors ($PC$) changes by one dollar the store manager’s evaluation also changes by 1.26, which is significant at $p = 0.0001$.

Although we did not formally state any hypotheses about the impact on SALES or GPROFIT due to unmeasured, omitted, or random effects on the store manager’s evaluation, it is interesting to examine these results. Examining these results we see that for SALES and GPROFIT, the manager receives a higher rating as these outputs rise ($PU$), and vice-versa. For example, as SALES due to unmeasured, omitted, and random effects rise by one dollar the manager receives 0.09 rating points and when GPROFIT rises by one dollar due to these effects the manager receives 0.30 rating points.

As hypothesized in H1B, target outcome knowledge or whether or not SALES or GPROFIT has been achieved significantly influences the manager’s evaluation. This illustrates the outcome effect is stubborn. For failure to meet the target sales the manager is penalized 11.56 points ($p = 0.0001$) and for failure to meet the gross profit target the penalty is 10.04 points ($p = 0.0001$).

**Discussion**

This field study provides evidence of the outcome effect in performance evaluations. Specifically, for the retail chain studied, we found that SALES and GPROFIT under the control of the manager influences the performance evaluation of that manager as expected, and that SALES and GPROFIT due to uncontrollable environmental factors do not influence the manager’s performance evaluation, also as expected. However, inconsistent with our expectations, the SALES

---

**Table 3**

Regression results for performance evaluation as a function of controllability of output factors and target outcome knowledge$^a$

$$R = B_0 + B_1(PE) + B_2(PC) + B_3(PD) + B_4(PU) + B_5T + e$$  

(T3.1)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Outcome measure = SALES</th>
<th></th>
<th>Outcome measure = GPROFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>$T$-stat.</td>
<td>$P$-value</td>
</tr>
<tr>
<td>Intercept</td>
<td>34.98</td>
<td>14.25</td>
<td>0.0001</td>
</tr>
<tr>
<td>$PE$</td>
<td>0.12</td>
<td>1.24</td>
<td>0.2154</td>
</tr>
<tr>
<td>$PC$</td>
<td>0.41</td>
<td>3.65</td>
<td>0.0003</td>
</tr>
<tr>
<td>$PD$</td>
<td>0.06</td>
<td>1.69</td>
<td>0.0501</td>
</tr>
<tr>
<td>$PU$</td>
<td>0.09</td>
<td>2.47</td>
<td>0.0145</td>
</tr>
<tr>
<td>$T$</td>
<td>$-11.56$</td>
<td>$-8.91$</td>
<td>0.0001</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.39</td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>Model $F$-value</td>
<td>26.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model $P$-value</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Variable definitions are as follows: $R$ = the supervisor’s valuation of the store manager’s performance, $PE$ = predicted economic outcomes due to environmental or uncontrollable factors, $PC$ = predicted economic outcome due to conditionally controllable factors, $PD$ = predicted economic outcomes due to controllable factors, $PU$ = all omitted and unmeasured determinants of store outcomes including random effects, $T$ = target outcome knowledge for either SALES or GPROFIT (coded as dummy variable; 1 for failure to meet the target and 0 for target achieved).
AND GPRPROFIT due to uncontrollable central management factors do influence the manager's evaluation. At first glance one would tend to consider this result as an anomaly since the store manager does not decide either the store location or the size of a store but is assigned to a store location and/or store size facility. However, as one might expect, managers are not assigned to stores randomly but based on obtaining a good fit between the store and the manager. When central management decides which manager to assign to which outlet, it is logical to assume that the assignment is conditional on the manager meeting some criteria (see e.g. Libby & Frederick, 1990), such as performance or experience, or else the manager would not be in charge of that particular store to begin with. Whatever that criteria may be, the key point is the store assignments are not random. These two evaluation factors (i.e. location and store size) have some information content about the manager’s prior ability to positively influence the probability distribution of store outcomes. Each of these evaluation factors is a source of information. What matters is the information content of the factors and whether or not the store manager can influence or control the information content of these factors, as opposed to the factors per se. Intuitively, store managers do exercise such an influence, which is why location and store size are, perhaps, significant explanatory variables of managers’ subjective performance evaluation. Incidentally, when the information content of a factor can be controlled or influenced, as opposed to the factor per se, it is the same as saying the factor is conditionally controllable since information content is linked to conditional controllability (Antle & Demski, 1988; Demski, 1994). Thus, store location and size may be considered as conditionally controllable factors by the store manager which may explain why they were given credit in the evaluation of these factors.

Despite the preceding findings this study did find strong evidence of an outcome effect when target outcome knowledge was taken into consideration as well as the controllable and uncontrollable determinants of segment outcome. Since this finding is consistent with prior results from experimental studies, the robustness of the outcome effect in subjective performance evaluations is evident.

In general, this study empirically demonstrated that the outcome effect is prevalent in an industry setting and is robust across the measure of outcomes. Kinney (1969) suggests that differences in performance of multi-unit businesses can be due to differences in the environmental factors confronting the multi-unit businesses and differences in the performance of the outlet managers (i.e. their managerial decision making ability). He suggests “the effects of these environmental factors should be extracted before evaluating the performance of the manager” (p. 44) because if the performance rating of the manager is influenced by uncontrollable environmental factors then the performance measurement and evaluation system is flawed. Further, the manager should be only held responsible for factors that are controllable by him or her. This research found that the multi-unit retail chain studied performed well in this regard. It was neither rewarding nor penalizing its managers for environmental factors that influence store performance but took into consideration the controllable factors. Also, it was giving the managers credit for unmeasured, random, or omitted variables. Hershey and Baron (1992) argue that while in an experiment the probability of various outcomes and other relevant information affecting outcome can be fully conveyed to the judge, “in more realistic cases, it is likely that the decision maker would have information relevant to the probability judgement that the judge cannot have”(p. 91). Some of the omitted variables could represent controllable factors we did not measure and are hard for the supervisors (judge) to have complete knowledge about (i.e. day to day operations such as cleanliness of store, employee relations, etc.) or would be difficult for the supervisor to know precisely. In other words, it is indeed possible the store manager has some influence over outcomes but the evaluator does not know the nature or the extent of that influence. In such situations, it is reasonable to give some credit to the segment manager for these outcomes.

As with any research effort, limitations exist, and the results of the present field study should be considered in light of those limitations. The most
serious weakness of a field study is its ex post facto character. Thus, statements of relations are weaker than they are in experimental research (Kerlinger, 1986). To complicate matters further, the field situation almost always has a plethora of variables. In an experiment, these variables can be largely controlled, but in a field study, they cannot be strictly controlled. However, they could be partially controlled if data on possible covariates is collected. Although this would still not represent the strict internal control of an experiment, it would be a step in the direction of greater control of extraneous factors. Another methodological weakness is the problem of precision in the measurement of field variables. In field studies, this problem is of course more acute than in experiments due to the greater complexity of field settings. In addition to the general problems associated with field study there are at least two which are more relevant for the current study. The research setting was in a retail chain. Therefore, identification of the controllable/uncontrollable factors in this study was industry-specific which limits the generalizability of the results. However, the research site is representative of other multi-unit organizations. The second limitation pertains to the data availability. The measures were constructed based on the data available and therefore an opportunity exists to develop better measures.

Future research should persevere in its endeavor to further investigate the subtleties of outcome effect and performance evaluation. Four possible avenues of research are suggested. First, research is needed on whether communicating the decision process, as well as the decision outcome, to the evaluator mitigates the outcome effect. For example, in addition to reporting the material price variance, management accountants may also report the assumptions that the DM had to make in order to estimate the initial budgeted price — quality of raw material, consumption patterns, market demands, supply sources, etc. This information would allow the evaluator to assess the validity of the critical assumptions and the uncertainty faced by the DM in budgeting the materials price. Second, one might also examine whether an education in debiasing strategies during training programs for supervisors reduces the outcome effects. Such an education would describe or warn the supervisors (see Fischhoff, 1982) of this potential problem in their evaluations of subordinates and may help to overcome, at least to some extent, their biased judgments. Third, one could assess if conducting the subjective performance of a manager either before or at various time lags (e.g. 1, 2, 4 or 6 weeks) after the objective target outcome knowledge on segment performance becomes available mitigates the outcome effect. Finally, since this research developed a field methodology for isolating outcome effects, we believe it would be prudent to study how various organizational factors or industry factors might influence the level of outcome effect.

This first attempt at uncovering outcome effects in a field study may encourage others to explore some of the preceding questions and also open the way for studies in other industry settings. This seems a natural step for building on the long tradition of high quality laboratory research that has been conducted which has addressed many issues in accounting and managerial decision making.

Acknowledgements

We would like to thank Margaret Boldt, Steve Butler, Uday Chandra, the two anonymous reviewers, and the workshop participants at the University of Oklahoma and Michigan State University for their comments on earlier versions of the paper.

References


