

Opening Pandora's Box:  
Studying the Accuracy of Managers' Perceptions

William H. Starbuck  
Leonard N. Stern School of Business  
New York University

John M. Mezas  
Leonard N. Stern School of Business  
New York University

1996

Working Paper Series  
Stern #IS-97-29

# Opening Pandora's box: Studying the accuracy of managers' perceptions

by

**William H. Starbuck**

and

**John M. Mezas**

**Published in the *Journal of Organizational Behavior*, 1996, 17(2): 99-117.**

## **Summary**

Both researchers and managers depend on the accuracy of managers' perceptions. Yet, few studies compare subjective with "objective" data, perhaps because it is very difficult to do well. These difficulties also muddy interpretations of results. On one hand, studies suggest that managers' perceptions may be very inaccurate. On the other hand, the observed errors in managerial perceptions may arise from research methods instead of managers.

Because perceptual data are so significant for both researchers and managers, researchers need to understand both the potential contaminants of perceptual research and the determinants of perceptual errors and biases. This article reviews studies of the accuracies of managers' perceptions, points out hazards in such research, and suggests various ways to improve studies of perceptions. The suggestions encompass improvements in gathering more valid subjective data, locating more appropriate "objective" data, finding appropriate respondents, and using statistical methods that provide accurate and reliable estimates with small samples.

## **Introduction**

This article both urges researchers to study the accuracies of perceptions and points out the difficulties of conducting such studies. Perceptual data are extremely important in many research and theoretical domains, and much has been written about perceptual biases. However, studies analyzing the validities of perceptual data are rare. For example, from 1988 to 1992, the *Journal of Organizational Behavior* published 249 articles. Only one of these compared perceptual with objective data and this comparison was more qualitative than quantitative (Jermier, Gaines, and McIntosh, 1989).

Because people's perceptions reflect their perceptual biases, observers' reports tell much about the observers: Sometimes the reports may even tell more about the observers than the reported phenomena (Starbuck and Milliken, 1988). Thus, it seems reasonable to ask questions about data that describe people's perceptions of their environments.

Such data are pervasive and often taken for granted. For example, "objective" government statistics, such as census data, often merely quantify subjective reports. Similarly, academic research objectifies perceptual data. Relying on the abstracts, we classified the 249 articles published by the *Journal of Organizational Behavior* from 1988 to 1992. 210 articles presented perceptual data, and 10 articles also presented "objective" data that merely aggregated perceptions. Other top journals probably have similar frequencies.

Yet, people's perceptions may often be very unrealistic. Many studies have reported finding large errors and biases in perceptions (Kahneman and Tversky, 1973; Kiesler and Sproull, 1982; Lant, Milliken and Batra, 1992). If perceptions incorporate large errors, then much published research describes errors or shared myths.

Erroneous perceptions also pose questions about some theoretical premises. Many theories assume that people's actions do

or should fit their environments -- for example, that managers do or should match organizations' properties to the organizations' environments. Of course, these theories also assume that managers base such actions on their perceptions of the organizations' environments. But what if managers' perceptions -- either of their organizations or of their organizations' environments or of both -- are very unrealistic? Or what if perceptions of organizations or organizations' environments vary greatly across managers?

Further, what sense should one make of prescriptions that researchers base on such data and premises -- say, prescriptions about the organizational properties that best suit certain environmental properties? For example, strategic planning may be as likely to reduce profits as to raise them. If strategies reflect accurate assessments and forecasts, everyone could work together to achieve difficult, but possible, goals, while ignoring irrelevant distractions. However, if planners have erroneous perceptions and make inaccurate forecasts, strategic planning could induce everyone to pursue the wrong goals, while overlooking unexpected opportunities.

Although these issues relate broadly to perceptual data obtained from people of all sorts, this article focuses on studies that use managers as primary data sources. Managers' perceptions form the bases for firms' actions and strategies as well as for much research. Many organizational theories assume that managers act purposely; to do this, managers must rely on their perceptions.

The next section reviews the few existing studies, which suggest that managers' perceptions do not correlate with "objective" measures. The ensuing sections suggest that studies comparing subjective with "objective" data may be rare because this topic is difficult to study. This difficulty may, indeed, explain why studies seem to have found inaccuracy.

### **Prior studies**

In a pioneering study, Lawrence and Lorsch (1967: 133) asked: "What types of organizations are most effective under different environmental conditions?" They inferred that firms have higher performances when managers align organizational properties with environmental properties. They said (1967: 134), "The most successful organizations tended to maintain states of differentiation and integration consistent with the diversity of the parts of the environment and the required interdependence of these parts." However, Lawrence and Lorsch obtained no "objective" measures of either organizational properties or environmental properties; all of their data were managers' perceptions.

Duncan (1972: 325) noted that perceptions of environmental properties likely vary with the individual differences among perceivers. "Some individuals may have a very high tolerance for ambiguity and uncertainty so they may perceive situations as less uncertain than others with lower tolerances." This implied that it should be important to distinguish subjective from "objective" measures, but he did not make this comparison.

Some researchers questioned Lawrence and Lorsch's questionnaire design. After attempting to replicate Lawrence and Lorsch's study, Tosi, Aldag and Storey (1973) began to doubt the reliability and validity of their measures of environmental uncertainty, and they decided to investigate these measures. Downey, Hellriegel and Slocum (1975) set out to improve on the work by Tosi et al. As a result, both groups asked middle and top managers to describe the stabilities of their markets. They then compared these perceptions with volatility indices calculated from the firms' financial reports and industry statistics, presuming that volatility correlates highly with uncertainty. The correlations between managers' perceptions and "objective" measures were all near zero and were negative more often than positive.

Tosi et al. (1973) analyzed data from 102 middle and top managers from 22 diverse firms in thirteen industries: The researchers calculated volatility indices from the firms' financial reports and industry statistics. These volatility indices correlated from -0.29 to +0.07 with the managers' perceptions of their environments' uncertainties. Tosi et al. interpreted these results as saying that Lawrence and Lorsch's "environmental subscale is not methodologically adequate."

Downey et al. (1975) obtained data from 51 heads of the divisions of one large conglomerate. They calculated three kinds of volatility indices and one index of competitiveness, and compared these with the managers' perceptions. The correlations ranged from -0.24 to +0.21. They too expressed doubts about the questionnaires, but they also expressed doubts about their own "objective" measures of environmental properties.

We accept the premise of Tosi et al. and Downey et al. that uncertainty should correlate with volatility, but their findings are subject to different inferences. In particular, these studies seem to say that, on average, managers' perceptions of their

environments' stabilities have no correlation with "objective" measures of those stabilities. Finding this interpretation very disquieting, we wondered whether similar incongruences occur with other environmental properties. However, Tosi et al. (1973) did not set out to compare managers' perceptions with "objective" measures; and although Downey et al. (1975) emphasized the comparison of subjective with "objective", they expressed doubts about the appropriateness of their "objective" measures. Thus, the low and erratic correlations they obtained might have arisen from poor "objective" measures, poor questionnaires, or inaccurate perceptions.

Whereas the preceding studies point to possible errors in managers' perceptions of their firms' environments, Payne and Pugh (1976) raised the possibility of similar errors in perceptions of organizational properties. They reviewed scores of studies in which researchers asked firms' members to characterize firms' structures and cultures. Their review suggests that most people see their firms inaccurately. They found: (a) Different members of a firm disagree so strongly with each other that it makes no sense to talk about an average perception. "Perceptual measures of each of the structural and climate variables have varied so much among themselves that mean scores were uninterpretable" (p. 1168). (b) Except for organizational size, members' perceptions of their firms correlate weakly with "objective" measures of their firms' properties. (c) Differences among members' perceptions of their firms' properties seem to correspond with their jobs and hierarchical statuses. For example, higher-status members generally have more favorable views of their organizations.

### *We looked and didn't find*

Intrigued by the foregoing studies, but wondering about their deficiencies, we decided to search for other studies that compared managers' perceptions with "objective" data. Because the validity of subjective data is so important to researchers, we expected to find quite a few relevant studies. Using combinations such as "percep\*" and "accura\*", "percep\*" and "consensus", "percep\*" and "objective", "objective" and "accura\*", "objective" and "consensus", "objective" and "subjective", "subjective" and "accura\*", and "subjective" and "consensus", we made a computer search of the Abstracted Business Index (ABI) for the 6.5 years from 1986 through mid 1992. To our astonishment, these searches turned up no studies that compared managers' perceptions with "objective" measures. Later searches covering 1988 through early 1994 also turned up no studies. One study did relate perceptions of uncertainty to experienced accuracies in prediction, but the subjects were students (Blomqvist, 1988).

We are not the only people who have searched for such studies without finding them. Sutcliffe (1994: 1360) reported that "surprisingly few empirical studies" compare managerial perceptions of environments with "objective" measures of environments. She pointed to two studies comparing subjective with "objective" measures. Neither study reported degrees of agreement between subjective and "objective" measures because they focused on correlations between perceptual errors and firm performance (Bourgeois, 1985; Dess and Keats, 1987). Also, both studies averaged the perceptual data across the top managers in each firm. Sutcliffe (1994) also averaged the perceptual data across several managers from each firm. Several studies have documented situations in which averaged perceptions were quite accurate even though almost all individuals perceived inaccurately (Dawes, 1977; Gordon, 1924; Starbuck and Bass, 1967; Zajonc, 1967). Thus, averaged perceptions likely overstate the accuracies of managers' perceptions.

Dess and Robinson (1984) also compared the perceptions of members of top management teams with "objective" measures of return on assets and sales growth. However, their "objective" measures were the perceptions of the chief executive officers, so their study actually compared two sets of perceptions.

It would be gross understatement to say that there have been few studies that focused on the differences between perceptions and "objective" data. There may be many studies in which researchers compared subjective and objective measures as minor ingredients in larger studies. However, our key-word searches did not detect such comparisons.

### **Why is it hard to study perceptual accuracies?**

Seeing the rarity of studies and their deficiencies, we set out to improve on what others had done. We formulated hypotheses, designed a study, gathered data, and analyzed them. These experiences illustrate why this subject is difficult to study. It is difficult, perhaps impossible, to design good questionnaires. It is difficult to obtain the right "objective" data. It is difficult to obtain enough appropriate respondents to make statistical analyses that are reliable.

These issues are not unique to studies of perception, of course: They lie at the heart of much research in organizational behavior. Thus, our observations probably apply more broadly even if these issues are more serious in the case of perceptual studies. Certainly, we have given more thought to these issues in the context of perceptual studies, and we use our study of

managers' perceptions to illustrate the issues.

The next subsection discusses questionnaire design. It calls attention to three types of questionnaire items and to the differences between researchers' worlds and managers' worlds. Then the ensuing subsections discuss the problems of getting good "objective" data and getting appropriate respondents. Although it is difficult to match perceptual data with suitable "objective" data, researchers can take several practical steps toward better "objective" data. Researchers need to obtain data from senior managers with differing amounts and kinds of experience. This implies small samples, but conventional squared-error regression requires large samples to produce reliable estimates. So the final, long subsections suggest using some alternative statistical techniques. In particular, one criterion that appears very promising is minimizing the number of outliers. This criterion yields more reliable coefficient estimates and more accurate estimates of effect sizes, and its advantage is greatest for small samples. Thus, it has special promise for studies of senior managers.

*Designing Good Questionnaires*

**Researchers may manufacture errors in managers' perceptions.** Statistical analyses require comparable data from each respondent, and this naturally led us to use questionnaires. Reviewing the questionnaires used by others disclosed troublesome issues that suggested some hypotheses for our study. For example, many of the questions that researchers have used to elicit managers' perceptions do more than they purport to do. The questions do not merely ask respondents to report facts; they ask for interpretations of facts. Some questions even invite respondents to invent facts where no facts can possibly exist.

Table 1 illustrates three types of questions. Interpretation seems to be intrinsic to questionnaire items that offer respondents choices among qualitative answers, whereas questions about facts usually ask for quantitative answers. For example, imagine two managers who give the same, accurate answer to the question: "By what percentage did your industry's sales grow during 1991?" Say, 10%. It is plausible that one manager might describe 10% growth as "dynamic" while the other might describe it as "static." The meanings of qualitative terms such as "dynamic" and "static" depend upon the managers' frames of reference. Someone from an industry that has been growing 3% per year might classify 10% as dynamic, while someone from an industry that has been growing 30% per year might regard 10% as static.

**Table 1. Three Types of Questions**

*Questions That Ask For Facts*

What is your firm's sales volume in dollars?

How many people does it employ?

How many times per year does your firm engage in formal strategic planning activities?

How quickly does your industry react after a leading firm takes a major action?

One	One	One	One	Longer
day	week	month	year	

*Questions That Ask For Interpretations*

Please circle the numbers that best describe your firm?

1	2	3	4	5
small sales volume				large sales volume
1	2	3	4	5
large work force				small work force

How often does your firm do formal strategic planning?

Very frequently	Frequently	Not very often	Rarely	Never
--------------------	------------	-------------------	--------	-------

How quickly does your industry react after a leading firm takes a major action?

Immediately	Quickly	Slowly	Very slowly
-------------	---------	--------	----------------

### *Questions Lacking Factual Answers*

How similar to each other are the firms in your industry segment?

Very similar	Somewhat similar	Somewhat diverse	Very diverse
-----------------	---------------------	---------------------	-----------------

What percentage would you estimate that strategizing has added to the performance of your firm during 1988-1991?

How frequently has your firm changed its strategies during 1988-1991?

More than once a year	Roughly once a year	Less than once a year	No changes recently
-----------------------------	---------------------------	-----------------------------	---------------------------

What portion of your firm's competitive actions adhere to pre-set strategies?

Almost all	Many	Some	Almost none
---------------	------	------	----------------

Some no-facts questions use ambiguous words - such as similar, often, and adhere. Although everyone understands the general idea of, say, adherence, people may disagree about how close something must be before one can say it adheres. Some no-facts questions call for multidimensional comparisons where no standards exist for aggregating across dimensions. A group of firms may be very similar along some dimensions, somewhat divergent along others, and very different along others; whether one classifies these firms as similar depends on how one weighs the dimensions. Some no-facts questions ask about social aggregates - such as a department or firm - and there are different standards for aggregating across people. Some norms say that each person should count equally; other norms say that the opinions of higher-level employees should count more. Some no-facts questions concern phenomena on which firms do not gather information systematically. Each employee has freedom to form an independent opinion based on any data the employee chooses. Some no-facts questions ask for comparisons with undefined standards. Thus, one could only say how much profit has been "added" by strategizing if one could say what profits would have been without strategizing.

What if researchers have found errors in managers' perceptions because of the ways in which they posed questions? That is, do questions having factual answers produce more accurate responses, and do the other two types of questions produce erratic responses that might appear inaccurate? In a study of assessments of uncertainty, Budescu, Weinberg, and Wallsten (1988) found greater variability in verbal responses than in numeric responses.

Other research studies and our experiences with questionnaires indicate that researchers face two serious challenges. First, researchers and managers operate in distinct cultures with different languages and different ideas that make communication between them very difficult. Second, when managers face very different situations, it becomes misleading to aggregate their questionnaire responses.

**Researchers and managers live in different worlds.** The differences between researchers and managers pose obvious issues of salience in that researchers may pose questions that have no meaning or very different meanings for managers (Labaw and Rapoport, 1980). For example, some research studies have asked managers to place their organizations on a scale ranging from "organic" to "mechanistic." Thinking this exemplifies researchers using terminology that managers do not understand, we tried this item on the executives in our sample. One manager wrote on the questionnaire form: "Do you mean do we grow it or do we make it?" Of course, this was a to-be-expected response to academic jargon.

However, our experience with another term -- industry -- indicates the difficulty of this problem. We asked managers, "In what industry does your business unit operate? Almost all of them gave answers that seemed appropriate to us. However, one person wrote: "My unit doesn't operate in 1 industry. Instead each of the staff focuses on a specific industry/segment." That is, this person was speaking of groups of customers rather than competing firms. Researchers consistently think of industries in terms of competing firms.

Using behaviorally anchored scales and non-verbal data might also help researchers to negotiate the ambiguities between themselves and respondents. Behaviorally anchored scales make ideas more specific. Conversely, drawing diagrams can help people describe ideas that words do not fit (Meyer, 1990).

Also, whereas managers get performance data from internal reports, researchers have shown preference for data that are conveniently available. American researchers nearly always study the "industries" defined by four-digit Standard Industrial Classification (SIC) codes. Only five of our 70 managers could equate their business units' competitive environments with SIC codes or labels approximating such codes. Two factors seem to explain managers' nonuse of SIC codes. First, few managers know what SIC codes are. Indeed, many respondents asked, "What is an SIC code?" Second, although a four-digit industry is the smallest SIC category, it is larger than the competitive environments that many managers find meaningful. In particular, many managers perceive their business units as competing in local or regional markets, whereas SIC codes describe the entire U. S. Other managers described their competitive environments with phrases such as "Jewish inground burial," "computer consulting for small businesses and not-for-profits," "entertainment advertising," "radio air personality," "EDI software," "architecture and construction of custom homes," "alkaline batteries," "emergency health care," and "industrial sorbents for oil-spill clean-up and chemical-spill clean-up."

The gap between managers and researchers is much more than linguistic. Managers operating single business units over time focus on properties that enable or indicate marginal changes. These include properties of business units such as policy changes and capital-equipment purchases, and properties of environments such as competitors' actions and current sales figures. Researchers conducting cross-sectional studies focus upon properties that distinguish organizations or environments. These include properties of business units such as firm size and market share, and environmental properties such as munificence and dynamism. Thus, managers may not be good sources of information about the properties that interest researchers, and researchers may not be studying phenomena that interest managers.

Cross-sectional research seems unlikely to provide information about the degrees to which managers act purposely or on the relations between managers' perceptions and their actions. To investigate such issues, researchers need to make longitudinal studies that attend to the properties that managers observe and control.

**Managers live in different worlds.** To researchers, managers' definitions of their worlds seem microscopic; but to managers, their worlds look dramatically different from each other. Consequently, as researchers make questionnaires more general, managers see the questions as less meaningful and they find it more difficult to produce meaningful answers. Also, very general questionnaires may overlook the information that managers use to observe and understand their worlds (Starbuck, 1981). Two managers may attend to very different stimuli even though they both operate in what researchers view as the same industry. Indeed, two managers may attend to very different stimuli even though they work in the same company. Thus, questionnaires may be inadequate as the sole means for gathering data about managers' perceptions.

One implication is that researchers need to adapt their information gathering to the language and ideas of specific managers. Pretesting can reduce the effects of academic jargon and can tailor items to specific industries or firms. However, dealing with individual differences among managers involves less reliance on written answers to questionnaires and more reliance on interactive exchanges in which researchers solicit information about how each respondent sees and describes her/his world. An optimal research strategy may not exist because tailoring questions to individual respondents increases validity at the expense of generalizability. Conversely, drawing more general inferences entails added risk that some respondents do not understand what researchers are asking.

### *Getting Good "Objective" Data*

Talking about the accuracies of perceptions implies that one has "objective" data against which to contrast perceptual data. The preceding section implicitly points to many difficulties in obtaining relevant "objective" data. For instance, divergent perceptions among co-workers make it very difficult to define "objective" answers to no-facts questions.

When dealing with no-facts questions, researchers have typically defined "objective" realities in terms of the average responses from many people. However, as Payne and Pugh (1976) remarked long ago, large variances across co-workers cast doubt on the idea that averages have meanings. For an average to summarize opinions meaningfully, one needs consensus (Brunsson, 1982).

Where perceptions are highly divergent, researchers might consider studying variations in consensus rather than accuracy of perceptions. In particular, consensus may be a more useful concept than accuracy for studies of organizational properties because so many of these are defined in subjective terms.

When dealing with fact questions, researchers have typically defined "objective" realities with conveniently available data. But, very often, such conveniently available data do not describe the environments that managers define as relevant to their organizations. Data concerning privately held firms are rarely public (Dess and Robinson, 1984). Boyd, Dess and Rasheed (1993) pointed out that archival measures are often at the wrong levels of aggregation or out-of-date. Indeed, archival data are usually not available during the year when perceptual data are gathered.

Researchers can make "objective" data more relevant to the environments that most managers see by obtaining data less aggregated than the four-digit SIC codes. For example, journals specializing in specific industries often report more specific data, and some consulting firms gather and sell specific data. Researchers can also make "objective" data more relevant by waiting until data become available concerning the time periods for which they gathered perceptual data. Where the relevance or accuracy of data is unclear, it may be helpful to compare data from different sources.

Researchers can also attend to the errors in convenient data. San Miguel (1977) found a thirty percent error rate in Compustat's reporting of R&D expenditures. These errors arose both from firms' reporting and from Compustat's processing. Rosenberg and Houglet (1974) examined the stock prices reported by Compustat and by the Center for Research in Security Prices at the University of Chicago. They (p. 1303) remarked, "There are a few large errors in both data bases, and these few errors are sufficient to change sharply the apparent nature of the data." Rosenberg and Houglet advised researchers to pinpoint errors by comparing data from different sources.

### *Getting Appropriate Respondents*

Researchers cannot expect studies of student subjects to apply directly to strategy formulation or organization design in firms. This is why existing studies obtained data from experienced, practicing managers, not students. Indeed, most of these studies got data from senior managers -- members of top-management teams or division heads. Many contemporary theories of strategic behavior talk about top management teams rather than individual managers. These theories imply a need to observe all or nearly all members of each top-management team.

At the individual level, work experience ought to have strong effects on perceptions, as managers with more experience or more relevant experience have had more opportunity to develop accurate perceptions. Thus, researchers should try to obtain data from managers with differing amounts and kinds of experience -- including some with extensive experience.

Of course, it is not that easy to persuade senior managers to participate in academic research projects. This challenge looms much larger if researchers need managers' cooperation on several occasions. Yet perceptual studies may require data from each respondent more than once, because respondents assume that they ought to see logical associations among the questions posed on one occasion. Indeed, respondents may invent logical associations that they did not imagine before coming across some grouped questions.

### *Getting Enough Respondents*



Because it is not easy to obtain appropriate respondents, it is important to use statistical techniques that do not require large samples.

**Some liabilities of conventional statistical techniques.** Almost all researchers use conventional statistical techniques, such as analysis of variance or squared-error regression, to analyze their findings. Unfortunately, very few studies obtain samples large enough to produce reliable estimates from such techniques.

Even when merely interpreting data at hand without making statements about new data they might gather, researchers need to beware of drawing unreliable inferences. All random samples include random errors that produce differences across samples. These random errors ought to have minimal influence on inferences; more reliable inferences are ones that depend less strongly on random errors. Thus, reliable inferences must apply implicitly to hypothetical new samples from the same populations.

To draw reliable inferences from moderately sized samples, researchers need to use something other than squared-error regression. The key issue is that squaring errors places great importance on the observations that lie far from the means. In regression, squared errors make the estimated coefficients depend strongly on the observations that are far from the regression lines. However, these outlying observations represent low-probability events that are unlikely to recur in other samples drawn from the same populations. Thus, sample idiosyncrasies have strong effects, and successive samples from the same population produce different coefficient estimates. Coefficients estimated from one sample tend to produce inaccurate predictions about another sample and about population parameters.

Sample idiosyncrasies also have strong effects because the criterion function -- the sum of the squared errors -- makes weak distinctions near its minimum. That is, the criterion function is flat and changes only a little when the estimated coefficients change a lot. As a result, coefficient estimates tend to be volatile.

Coefficient estimates have low reliability with small samples and become more reliable as sample sizes increase. Since a small sample contains few outliers, each outlier exerts strong influence and the peculiarities of specific outliers have great importance. Large samples have many outliers that tend to offset each other, so each outlier should exert weak influence and the peculiarities of specific outliers should have little importance.

How large should a sample be before one should rely on squared-error statistics? Several researchers have explored this issue in the context of admission and employment tests. There is an old tradition of evaluating applicants for jobs or for college admission by checking off their characteristics on lists. Evaluators then added up the numbers of checkmarks to determine applicants' suitability, which implicitly gave each item equal weight. For example, the first college entrance exams assigned equal weight to each question. The items on entrance exams or employment forms did not come from statistical studies but from a priori assumptions.

Then some psychometricians began to advocate using squared-error regression to assign weights to items (Perloff, 1951). The psychometricians argued that regression would assign higher weights to more important items and would eliminate redundant or uninformative items. Regression weights were said to minimize prediction errors. To the surprise of many, scores computed with regression-derived weights correlated less highly with students' or employees' actual performances than did scores generated by equally weighted a priori items (Boyce, 1955; Lawshe and Schucker, 1959; Wesman and Bennett, 1959).

Schmidt, Claudy, and Dorans and Drasgow used computer simulations to investigate this phenomenon. They assumed an ideal situation: perfect normal distributions and independent variables with no measurement errors. Schmidt (1971) examined sample sizes ranging from 25 to 1000. With ten independent variables, regression was inferior to equal weights unless there were more than 100-200 observations. Furthermore, even with samples of 1000, regression coefficients were only slightly superior to equal weights. Claudy (1972) examined sample sizes ranging from 20 to 160 and one to five independent variables. With independent variables that intercorrelated between 0 and 0.4, regression was inferior to equal weights for all sample sizes. Even with the largest samples (160), regression was either inferior or only slightly superior to equal weights for over half of the tested populations. Dorans and Drasgow (1978) compared six methods, and found that equal weighting outscored the other five methods when making predictions about new samples from the same population. Equal weighting had as high cross-sample validity for small samples as for large ones, whereas the other methods had lower validities for smaller samples. Regression drew the least reliable inferences from small samples and the most reliable inferences from large samples.

Einhorn and Hogarth (1975) used algebra to compare regression with equal weights, and they introduced another factor -- the magnitude of the multiple correlation coefficient ( $R^2$ ). They concluded: (a) Over wide ranges of sample sizes and numbers of independent variables, there is little difference between regression weights and equal weights. (b) Equal weights are more reliable than regression weights when samples are small, when multiple correlations are not high, or when independent variables intercorrelate. With ten independent variables and multiple correlation coefficients below 0.5, samples might have to exceed 400 before regression weights become more reliable than equal weights. (c) Because they assumed perfect normal distributions, their calculations underestimated the relative advantages of equal weights relative to regression weights.

**Simulations of squared-error regression with a perception study.** Our initial reaction to the foregoing studies was to wonder what they implied for our study of managers' perceptions. How many managers would we need in order for our statistical calculations to be more reliable than plausible a priori assumptions? The studies cited above led us to expect that we would need data from at least 150, perhaps 200 managers. But our data did not look exactly like a multivariate normal distribution, so we speculated that we might need data from more than 200 managers.

We decided to simulate our statistical problem. Our simulations generated hypothetical data that embody assumptions about how managers behave. These assumptions are partly based on data from 70 managers in two Executive MBA programs and are partly speculative. To simulate, we had to augment our data by making assumptions. Although we regard these as realistic assumptions, they came from introspection.

In particular, the simulations allow for the additive effects of five independent variables. They assume that perceptual error is smaller:

- if a manager's functional area relates to the perceived variable,
- if a manager has more total work experience,
- if a manager has more experience in the relevant industry environment,
- if a firm emphasizes informal planning and open communication, and
- if a firm plans more frequently.

Table 2 lists these variables together with descriptive statistics. The frequency distributions match those characterizing the Executive MBAs. Note that these empirical distributions do not closely resemble normal distributions. One variable is binary with a 25% chance of equaling 1. Two variables take integer values from 0 to 4. Two variables take integer values from 1 to 13, with means toward the lower end of this range.

Besides the perceptual errors in the independent variables, we added errors to the dependent variable to make the multiple correlation coefficients approximate 0.4. All these errors were normally distributed. Since existing studies have produced multiple correlations between 0.1 and 0.2, our simulations may understate the sample sizes needed for reliable regression calculations. However, we were feeling optimistic about the correlations our study would produce.

---

**Table 2. Variables Hypothesized to Reduce Perceptual Errors**

Variable	Range	True Mean	True Std Dev	Std Dev of Perceptual and Measurement Error (%)
Is manager's functional area related to topic?	1=related, 0=unrelated	0.25	0.43	17
Manager's total work experience	1 to 13 years	4.24	2.79	6
Manager's experience in relevant industry environment	Years: focal <= total	1.67	2.06	17
Informal planning and open communication	4=much, 0=none	2.00	1.41	40
Frequency of formal planning	0 to 4 times per year	1.98	1.17	23

Correlation between planning informality and frequency varies from -.25 to -.30.

The key control condition in these simulation experiments is the sample size. The computer generated N true values of the independent and dependent variables, added errors, ran a squared-error regression on the sample, and saved the estimated coefficients. Next, the computer generated a new sample of N observations, and calculated estimated values of the dependent variable by two methods:

1. The standardized value of the dependent variable equals (1<sup>5</sup>) times the sum of the standardized values of the independent variables.
2. The dependent variable equals the linear function estimated by squared-error regression from the preceding sample.

The computer repeated this process 50 times for each sample size -- 20, 50, 100, 200, 500, and 1000.

Setting the coefficient for the equal weights to (1<sup>5</sup>) made the standard deviation of the dependent variable approximately equal to the standard deviation of the sum of the independent variables. If the five independent variables were all uncorrelated with each other, the standard deviation of their sum would be "5. Since some of the independent variables correlated with each other, "5 is not entirely accurate, and researchers intending to use only equal weighting would doubtless estimate this coefficient from data. However, we were merely seeking to provide a benchmark, not to make equal weighting work as well as possible.

**Figure 1** compares the fit with regression to the fit with equal weights. Fit is measured by the covariance of the estimated and actual values of the dependent variable, which is analogous to the multiple correlation coefficient. Where the ratio exceeds 1, regression made more reliable estimates than did equal weights. Conversely, where the ratio falls below 1, equal weights made more reliable estimates than regression. The average ratio lies below 1 for all sample sizes. That is, equal weights produced more reliable estimates on average even with samples as large as 1000.

As well as the average values of the fits produced, **Figure 1** shows approximate 90% confidence intervals for the fits. That is, roughly 5% of the fits should fall above the line designated 95% and roughly 5% should fall below the 5% line.

Both the "most reliable" regression estimates and the "least reliable" ones occurred with small samples. The words "most reliable" and "least reliable" are in quotes because these results arise from the random errors in consecutive samples. Small samples increase the likelihood of two consecutive samples being unusually similar or dissimilar. When the consecutive

samples are unusually similar, regression produces apparently "reliable" estimates; conversely, when the consecutive samples are dissimilar, regression produces "unreliable" estimates.

Figure 2 graphs the maximum errors in estimated effect sizes -- the estimated changes in the dependent variable resulting from one-standard-deviation changes in the independent variables. Since the data arise from a computer, the true values of the population parameters and the coefficients in the formulas are known. One can calculate the percentage errors in standard deviations and coefficient estimates, and therefore the errors in effect sizes. Each datum going into Figure 2 is the maximum error associated with each simulation run; that is, these are the worst estimates of effect sizes. The average value of these maximum errors exceeds 100% for samples of 200 or less, and the 95% value exceeds 100% even with samples of 1000.

The largest errors in estimated effect sizes occur where managers' perceptual errors have large standard deviations and where independent variables have small coefficients. In fact, the ratio

$$\frac{\text{Standard deviation of perceptual errors}}{\text{Absolute value of coefficient}}$$

correlates 0.9 with the frequencies with which the maximum errors in estimated effect sizes occur with specific independent variables.

In summary, squared-error regression has a significant likelihood of yielding large errors in estimated effect sizes for all the sample sizes that researchers typically obtain. With sample sizes below 1000, estimates about dependent variables based upon regression calculations are less reliable than the estimates researchers could probably make solely on a priori grounds. Similar observations apply to other squared-error statistics such as analysis of variance. For evaluating the effects of independent variables, squared-error statistics are not reliable.

Of course, two factors affect the relative effectiveness of squared-error statistics versus a priori assumptions. First, squared-error statistics work better when samples are large and when data closely approximate multivariate normal distributions. Second, a priori assumptions work better where theorists better understand the phenomena they are studying.

The foregoing results both surprised and disappointed us. Instead of finding out whether we would need data from 150 or 200 managers, we had discovered that even 1000 managers would not ensure reliable analyses. Other researchers might be able to obtain samples large enough to make squared-error statistics reliable -- perhaps 5000 or 10000 respondents. However, we could not see any way to obtain that many appropriate respondents for our study. Seeking out a large sample would degrade the quality of our data by forcing us to include less appropriate respondents.

**Some unconventional statistical techniques.** We decided to investigate the possibility that other estimation criteria might offer higher reliability than squared errors: There are indeed criteria that promise better results.

Although even some who teach the subject are unfamiliar with the limitations of squared-error regression, these limitations are rather well-known to statisticians who specialize in regression analysis. Statistics journals have published quite a few articles contrasting squared-error regression with other criteria that sometimes prove superior. These alternative criteria include: least-squares regression with trimmed data, trimmed quantile estimation, quadratic mode regression, Tukey's three-group median procedure, minimizing the median of the squared residuals, minimizing the sum of the absolute-percentage errors, minimizing the sum of the absolute errors, and minimizing the percentage of outliers.

We decided to try the last two of these alternative criteria. The model underlying our simulations suggested that absolute errors would be a more appropriate criterion than absolute percentage errors. As Figure 3 illustrates, absolute errors increase less explosively than the squares of the errors. As a result, coefficient estimates that minimize the sum of the absolute errors tend to make more reliable predictions about new samples. For example, in a study of price expectations, Wiginton (1972) compared predictions based upon squared-error regression with predictions based on minimizing absolute errors. He found that the absolute-error criterion yielded more accurate forecasts even when he used the sums of squared errors to measure the accuracies.

Makridakis et al. (1987) investigated minimizing the observed percentage of outliers. This criterion seeks to explain the

data in a way that minimizes the numbers of observations that appear anomalous. It ignores the exact numeric values of the outlying observations; all outliers have the same weight. It also ignores the numeric values of observations that are not outliers. Since random errors affect the specific numeric values, this nonparametric criterion tends to be more stable across different samples from the same population (Pant and Starbuck, 1990).

**Figure 4** graphs three ratios: (a) the fit with regression versus the fit with equal weights, (b) the fit minimizing the sum of absolute errors versus the fit with equal weights, and (c) the fit minimizing the number of outliers versus the fit with equal weights. In each case, fit is measured by the optimized property. Where the ratios exceed 1, the fitting criterion made more accurate estimates than did equal weights. Conversely, where the ratio falls below 1, equal weights made more accurate estimates than the fitting technique.

The line corresponding to squared-error regression is the same one shown in Figure 1. This line falls below 1 for all sample sizes, thus indicating that regression is less reliable than equal weights. By contrast, the line corresponding to the sum of absolute errors exceeds 1 for all sample sizes, indicating that this criterion is more reliable than equal weights. The line corresponding to the number of outliers lies far above the others, indicating that this criterion produces significantly more reliable fits than the others.

**Figure 5** graphs the maximum errors in estimated effect sizes. The upper line, which corresponds to squared-error regression, is the same as the middle line in Figure 2. The middle line in Figure 5 corresponds to fits minimizing the sum of absolute errors, and the lowest line corresponds to fits minimizing the number of outliers. Remarkably, fits based on the number of outliers produce nearly the same maximum errors in effect sizes for all sample sizes. That is, although this criterion is superior for all sample sizes, its advantage is greatest for small samples.

**Two recommendations.** First, researchers ought to use simulations when designing studies. Our simulations not only gave us insights into the requirements of our project, they gave us surprising insights. Other researchers are likely to have similar surprises. Second, researchers should seriously consider using techniques other than squared-error regression, analysis of variance, LISREL, and other squared-error techniques. Many alternative techniques deserve consideration, any of which might be best for a specific study. However, our simulations indicate that minimizing the number of outliers offers great promise. This criterion may be especially suitable for small-sample studies, such as studies of senior managers.

## Conclusion

Perceptual data play extremely important roles in both studies of and theories about managerial behavior and in the behavior itself. Researchers use perceptual data to define "objective" situations. Organization and strategy theorists talk about matching organizations or strategies with environments. Managers act on their perceptions.

Yet, a few studies suggest that on average, managers' perceptions of their environments do not correlate with "objective" measures of those environments. Studies also suggest that many co-workers disagree strongly about the properties of their organizations. Thus, perceptions may be very inaccurate. To find out whether this is so, researchers need to investigate the determinants of larger or smaller errors and of different biases.

Nevertheless, surprisingly little research focuses on the accuracies of perceptions. Studies comparing subjective with "objective" data may be rare because it is so difficult to design good questionnaires, to obtain good "objective" data, and to obtain enough suitable respondents. Until researchers address these issues effectively, it remains debatable whether the observed errors in managerial perceptions arise from managers or from research methods.

Not only is it difficult to bridge the cultural gaps between managers and researchers, but different managers see very different worlds. As a result, written questionnaires tend to introduce errors through misunderstandings between researchers and respondents. Although it may be impossible to define an optimal data-gathering strategy, researchers can use a variety of techniques to improve communication between themselves and managers. For example, behaviorally anchored scales combat ambiguity, as does tailoring questions to specific industries or firms. Numeric responses may be more accurate than verbal categories, and nonverbal data allow respondents to express ideas that they cannot easily put into words. Interactive exchanges allow researchers to solicit information about how respondents describe their worlds and to attend to individual differences among managers.

Researchers can use "objective" data that are more relevant to managers' perceptions of their firms' environments by

searching out less-aggregated data and by waiting until appropriate data become available. Where the relevance or accuracy of data is unclear, it may be helpful to compare data from different sources and to investigate the errors in archival data. Where "objective" data do not exist and managers hold highly divergent perceptions, researchers might better study variations in consensus rather than accuracy of perceptions.

For research findings to apply to real-life strategizing or organization design, researchers need data from experienced, practicing managers -- especially senior managers and especially the members of top-management teams. To keep respondents from seeing logical interactions among simultaneously presented questions, researchers need to gather data from each respondent more than once. As a result, past studies have had and future studies will have small samples.

The existing studies have not had samples large enough to produce reliable estimates from conventional statistical techniques such as squared-error regression. To produce reliable estimates, these techniques require at least 150-200 degrees of freedom, often much more. Simulations based on our study of managers' perceptions indicated that even samples of 1000 would be too small for squared-error regression to produce accurate estimates of effect sizes.

Because of the disadvantages of conventional techniques, researchers should seriously consider using techniques other than squared-error regression and other squared-error techniques. In particular, functions that minimize the numbers of outliers hold remarkable promise. They appear to be especially useful for studies with small samples.

### *Out of the box and into the future*

Researchers have much to gain and nothing to lose by studying perceptual accuracies. First, research methods can be improved. If we can see opportunities for improvement today despite the paucity of studies, more studies are likely to disclose more opportunities for improvement tomorrow. Second, if today's apparent inaccuracies in managers' perceptions are largely due to methodological defects, and better studies show managers' perceptions to be quite accurate, theorists will have a stronger foundation for theories about managerial behavior. Third, if managers' perceptions are moderately inaccurate, we can study systematic biases and seek ways to correct for these. Fourth, if managers' perceptions are very inaccurate, we can pursue at least two fascinating puzzles.

**Puzzle One:** How can people act effectively despite very inaccurate perceptions? Perhaps, some perceptions are much more important. Perhaps, feedback about errors stimulates adjustments. Perhaps, slack resources insulate people and organizations from harsh consequences (Simon, 1956).

**Puzzle Two:** What are the optimal degrees and kinds of perceptual distortion? Baumeister (1989) argued that optimal psychological functioning involves moderate distortions in perceptions of self and world. For example, entrepreneurs might gain motivation from seeing opportunities as brighter than they really are. Similarly, at an organizational level, people's willingness to implement a policy enthusiastically depends on their conviction that it will produce excellent results (Brunsson, 1982).

### **Footnote**

This article benefits from the helpful suggestions of Cary Cooper, Lisette Losada, Stephen Mezas, Frances Milliken, and Elizabeth Morrison.

### **References**

- Baumeister, R. F. (1989). 'Self-illusions: When are they adaptive?' *Journal of Social and Clinical Psychology*, **8**: 176-189.
- Blomqvist, H. C. (1988). 'Uncertainty and Predictive Accuracy: An Empirical Study,' *Journal of Economic Psychology*, **9**: 525-532.
- Bourgeois, L. J. (1985). 'Strategic goals, perceived uncertainty, and economic performance in volatile environments,' *Academy of Management Journal*, **28**: 548-573.
- Boyce, J. E. (1955). Comparison of methods of combining scores to predict academic success in a cooperative engineering

program. Unpublished Doctoral dissertation, Purdue University, Lafayette, IN.

Boyd, B. K., Dess, G. G., and Rasheed, A. M. A. (1993). 'Divergence between archival and perceptual measures of the environment: Causes and consequences,' *Academy of Management Review*, **18**: 204-226.

Brunsson, N. (1982). 'The irrationality of action and action rationality: Decisions, ideologies, and organizational actions,' *Journal of Management Studies*, **19**: 29-44

Budescu, D. V., Weinberg, S., and Wallsten, T. S. (1988). 'Decisions based on numerically and verbally expressed uncertainties,' *Journal of Experimental Psychology: Human Perception and Performance*, **14**: 281-294.

Claudy, J. G. (1972). 'A comparison of five variable weighting procedures,' *Educational and Psychological Measurement*, **32**: 311-322.

Dawes, R. M. (1977). 'Suppose we measured height with rating scales instead of rulers,' *Applied Psychological Measurement*, **1**: 267-273.

Dess, G. G., and Keats, B. W. (1987). 'Environmental assessment and organizational performance: An exploratory field study,' *Academy of Management, Proceedings of the Annual Meeting*, 1987: pp. 21-25.

Dess, G. G., and Robinson, R. B., Jr. (1984). 'Measuring organizational performance in the absence of objective measures: The case of the privately-held firm and conglomerate business unit,' *Strategic Management Journal*, **5**: 265-273.

Dorans, N., and Drasgow, F. (1978). 'Alternative weighting schemes for linear prediction,' *Organizational Behavior and Human Performance*, **21**: 316-345.

Downey, H. K., Hellriegel, D., and Slocum, J. W. Jr. (1975). 'Environmental uncertainty: the construct and its application,' *Administrative Science Quarterly*, **20**: 613-629.

Duncan, R. B. (1972). 'Characteristics of organizational environments and perceived environmental uncertainty,' *Administrative Science Quarterly*, **17**: 313-327.

Einhorn, H. J. and Hogarth, R. M. (1975). 'Unit weighting schemes for decision making,' *Organizational Behavior and Human Performance*, **13**: 171-192.

Gordon, K. (1924). 'Group judgments in the field of lifted weights,' *Journal of Experimental Psychology*, **7**: 398-400.

Jermier, J. M., Gaines, J., and McIntosh, N. J. (1989). 'Reactions to physically dangerous work: A conceptual and empirical analysis,' *Journal of Organizational Behavior*, **10**: 15-33.

Kahneman, D., and Tversky, A. (1973). 'On the psychology of prediction,' in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, pp. 48-68.

Kiesler, S., and Sproull, L. (1982). 'Managerial responses to changing environments: Perspectives on problem sensing from social cognition,' *Administrative Science Quarterly*, **27**: 548-570.

Labaw, P. J., and Rapoport, M. A. (1980). *Advanced Questionnaire Design*. Cambridge, MA: Abt Books.

Lant, T. K., Milliken, F. J., and Batra, B. (1992). 'The role of managerial learning and interpretation in strategic persistence and reorientation: An empirical exploration,' *Strategic Management Journal*, **13**: 585-608.

Lawrence, P. R., and Lorsch, J. W. (1967). *Organization and Environment*. Boston: Graduate School of Business

Administration, Harvard University.

Lawshe, C. H., and Schucker, R. E. (1959). 'The relative efficiency of four test weighting methods in multiple prediction,' *Educational and Psychological Measurement*, **19**: 103-114.

Makridakis, S., Hibon, M., Lusk, E., and Belhadjali, M. (1987). 'Confidence intervals: An empirical investigation of the series in the M-competition,' *International Journal of Forecasting*, **3**: 489-508.

Meyer, A. D. (1990). 'Visual data in organizational research.' *Organization Science*, **2**: 218-236.

Pant, P. N., and Starbuck, W. H. (1990). 'Innocents in the forest: Forecasting and research methods.' *Journal of Management*, **16**: 433-460.

Payne, R. L., and Pugh, D. S. (1976). 'Organizational structure and climate,' in M. D. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, pp. 1125-1173.

Perloff, R. (1951). Using trend-fitting predictor weights to improve cross-validation. Unpublished Doctoral dissertation, The Ohio State University, Columbus.

Rosenberg, B. and Houglet, M. (1974). 'Error rates in CRSP and Compustat data bases and their implications,' *Journal of Finance*, **29**: 1303-1310.

San Miguel, Joseph G. (1977). 'The reliability of R&D data in Compustat and 10-K reports.' *Accounting Review*, **52**: 638-641.

Schmidt, F. L. (1971). 'The relative efficiency of regression and simple unit predictor weights in applied differential psychology,' *Educational and Psychological Measurement*, **31**: 699-714.

Simon, H. A. (1956). 'Rational choice and the structure of the environment,' *Psychological Review*, **63**: 129-138.

Starbuck, W. H., and Bass, F. M. (1967). 'An experimental study of risk-taking and the value of information in a new product context,' *Journal of Business*, **40**: 155-165.

Starbuck, W. H. (1981). 'A trip to view the elephants and rattlesnakes in the garden of Aston,' in A. H. Van de Ven and W. F. Joyce (eds.), *Perspectives on Organization Design and Behavior*. New York: Wiley-Interscience, pp. 167-198.

Starbuck, W. H., and Milliken, F. J. (1988). 'Executives' perceptual filters: What they notice and how they make sense,' in D. C. Hambrick (ed.), *The Executive Effect: Concepts and Methods for Studying Top Managers*. Greenwich, CT: JAI Press, pp. 35-65.

Sutcliffe, K. (1994). 'What executives notice: Accurate perceptions in top management teams,' *Academy of Management Journal*, **37**: 1360-1378.

Tosi, H., Aldag, R., and Storey, R. (1973). 'On the measurement of the environment: an assessment of the Lawrence and Lorsch environmental uncertainty subscale,' *Administrative Science Quarterly*, **18**: 27-36.

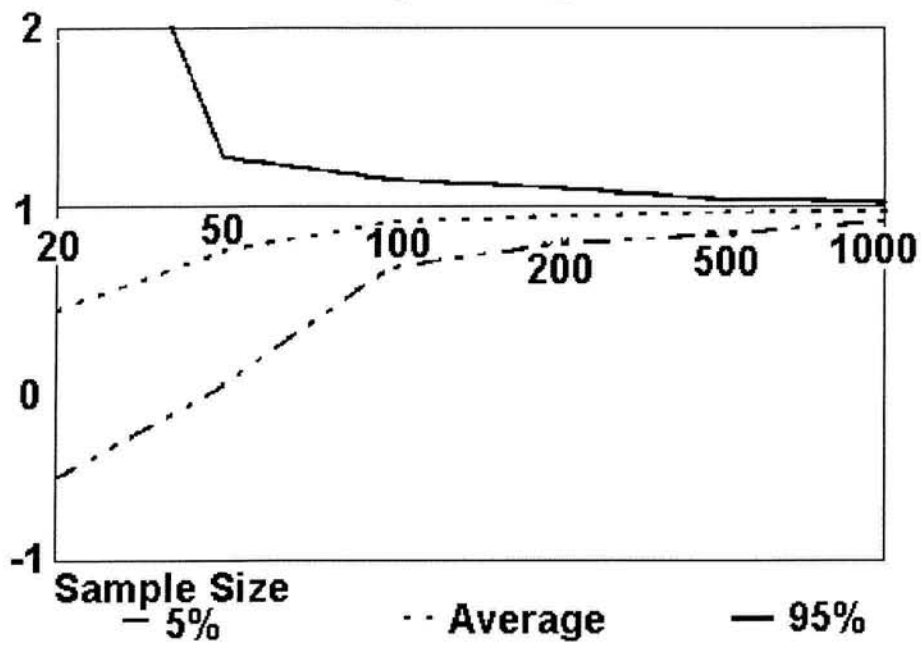
Wesman, A. G., and Bennett, G. K. (1959). 'Multiple regression vs. simple addition of scores in prediction of college grades,' *Educational and Psychological Measurement*, **19**: 243-246.

Wiginton, J. C. (1972). 'MSAE estimation: An alternative approach to regression analysis for economic forecasting applications,' *Applied Economics*, **4**: 11-21.

Zajonc, R. B. (1967). *Social Psychology: An Experimental Approach*. Belmont, CA: Brooks/Cole.



**Figure 1. Ratio of Fit with Regression to Fit with Equal Weights**



**Figure 2. Maximum Errors in Effect Sizes with Regression Estimates**

