Using self-report questionnaires in OB research: a comment on the use of a controversial method

PAUL E. SPECTOR

University of South Florida, U.S.A.

Introduction

Most studies of organizational behavior rely on reports by people about the variables of interest. The use of one type of report, the self-report where people are asked about their own jobs and their reactions to their jobs, has become the target of criticism by both article authors and reviewers. Many researchers are skeptical about results that come from questionnaires that ask people to report about themselves and their jobs. Skepticism is reflected in reviewers’ comments when a paper is dismissed with the often prejudiced and unthoughtful charge that method variance or monomethod bias has produced the observed correlations rather than the constructs of interest. It is less likely that this charge will be raised when people report about others, as with assessment center ratings or performance ratings, even though a single method is used here as well.

There are good reasons to be cautious in the use of self-report questionnaires, but reasons for caution are every bit as important for other methodologies as well. For example, Fried, Rowland and Ferris (1984) outlined the many problems that occur with the use of objective physiological measures. There have been serious questions raised about the validity of dimension ratings in assessment centers (e.g. Brannick, Michaels and Baker, 1989; Robertson, Gratton and Sharples, 1987). Frequent discussions of the problems with self-reports can be found throughout the OB literature (e.g. Brief, Burke, George, Robinson and Webster, 1988; Spector, 1992). Indeed self-reports have been used too frequently to address research questions that they are unable to adequately answer. It seems important, however, that we also consider the appropriate uses and values of self-reports. This paper will be concerned with how self-reports can and cannot be used to help us understand organizational phenomena. It will deal with both issues of construct validity and issues of appropriate inferences that can be made from studies using self-report questionnaires.

* Requests for reprints should be addressed to Paul E. Spector, Department of Psychology, University of South Florida, Tampa, FL 33620. The author thanks Michael T. Brannick, and Paul T. Van Katwyk for their helpful comments about this paper.

CCC 0894–3796/94/050385–08
© 1994 by John Wiley & Sons, Ltd.

Accepted 18 February 1994
What do self-report measures indicate?

The first important issue for this paper concerns what constructs or traits self-report measures can be considered to represent. In other words how can we reasonably interpret the meaning of an observed score on a variable of interest? There seems to have been relatively little criticism in the literature of self-reports as measures of people’s feelings about and perceptions of work. I suspect that most researchers and reviewers would accept that measures of job satisfaction and other affective reactions to work are valid indicators of people’s feelings about work, and that measures of the job environment probably do a reasonably good job of reflecting people’s perceptions of their jobs. Criticism of self-reports have focused mainly on their use as indicators of the objective job environment (e.g. Frese and Zapf, 1988; Spector, 1992).

Spector and Brannick (in press) provided a framework with which to interpret the meaning of a measured variable. According to their scheme, the variance of every measured variable can be partitioned into three components — trait, method, and error. Trait variance is the variance attributable to the construct of interest, and is dependent upon the researcher’s interpretation of the measured variable. The trait component of a self-report scale will be a function of whether one considers it to represent the objective environment or the respondent’s perception of the environment. The proportion of variance in the observed variable accounted for by the trait component will certainly be smaller in the former situation.

Method variance is variance produced by all other systematic influences on the measured variable. Again, the interpretation of the measured variable determines the extent to which variance is attributed to method. According to Spector and Brannick, variance attributable to method is the byproduct of both the method of measurement and the intended traits. Influences on a measured variable are sources of method variance that affect the assessment of particular traits when measured in a certain way. For example, answers to personally sensitive questions are likely to be influenced by social desirability while answers to less sensitive questions are not likely to be influenced, even though the format of the question or the ‘method’ is the same (Spector and Brannick, in press). This is somewhat different from the more popular notion that method variance is inherent in a particular method, such as a questionnaire, and that it influences all traits assessed with that method.

Error variance is produced by random error of measurement, which is due to nonsystematic influences on measured variables. It is reflected in the unreliability of measures. Every measured variable contains all three components, although the relative magnitudes of the three can vary considerably depending upon the interpretation of the measure, the method used, and the reliability of the measure.

We consider valid any measure for which there is reasonable evidence that supports our inference about it and our interpretation of what it represents. However, even measures that are considered valid indicators of an intended trait can be influenced by many other variables. For example, measures of some variables and not others may be affected by social desirability bias (Moorman and Podsakoff, 1992). Furthermore, there is mounting evidence that reports of job conditions are affected by many things, including respondent attitudes, cognitive processes, mood, and personality (see Spector, 1992).

Spector (1992) reviewed the existing evidence for the validity of self-report measures of job conditions. He estimated that about 10 per cent to 20 per cent of the variance of these measures represents the objective environment or intended trait. This estimate was based on two types of studies — those that provided the convergent validity correlations of different types of job condition measures, and experiments in which job conditions were manipulated.

Even those job condition measures that are considered valid, however, can contain large
method components. The measures considered by Spector had reliability levels that would result in about 10 to 30 per cent error variance. (This estimate is calculated as 1 minus reliability.) The effect sizes from the validity studies summarized by Spector (1992) suggest that about 10 to 20 per cent of the variance is attributable to trait. This leaves about 50 to 80 per cent method variance produced by sources other than that intended.

Should a measure that has only 10 to 20 per cent trait variance be considered valid? This question has no objective answer so the reader must decide. Two issues should be considered, however. First, is this level of trait variance worse than that of other types of measures? This question is difficult to answer because we do not have good estimates of the amount of trait variance in different measures. In one domain, performance measurement, different methods have been found to have only modest convergent validities, suggesting that their level of validity is not unlike that of job condition measures. If 10 to 20 per cent is inadequate, we will probably eliminate many of the measures used in the OB domain.

A second and more important issue is whether these measures can be useful. Here we must ask what self-report measure of job conditions might be able to tell us. Since they have been shown to be sensitive to job changes, it seems reasonable to use self-reports as outcome variables in studies where job conditions have changed or where job conditions are assessed objectively. An example of this sort of study was Griffin (1991) where job condition reports were used to determine if employees perceived changes that had taken place in their jobs. This longitudinal study using self-reports allowed Griffin to conclude that employees perceived the changes over a four-year span but their satisfaction was only affected for a short period. These results suggest that employee reactions to job changes can be more complex than suggested by many models of job conditions.

On the other hand, self-report job condition measures cannot be very helpful as independent variables where inferences are to be made about their effects on other self-report variables. When self-report variables are correlated with one another, it is difficult to know if it is the trait or method components that are responsible for the observed correlation. Because there can be many influences on a measured variable, evidence for construct validity is not sufficient to allow us to assume that correlations of the measure with other variables represent the relation of the intended trait with other variables. The method components in organizational measures are sufficiently large to account for observed correlations among variables in most studies. This produces excessive uncertainty in drawing inferences about the causes of correlations in OB research, the second major issue to be addressed by this paper.

**Drawing inferences from correlations among self-reports**

The reasonableness of using self-reports depends upon the purpose of the study. What inferences can we reasonably make about relations among constructs when all our variables are assessed with self-reports? In part the answer depends upon the design of the study. Objections to self-report studies are most strongly directed to those with cross-sectional designs where all data are collected at one point in time. Longitudinal studies allow for more confident conclusions about causal relations, which are difficult with cross-sectional designs, regardless of measurement method. Using multiple methods or sources of data (such as observers or supervisors) can also expand the confidence with which conclusions can be drawn from a set of data.

However, even longitudinal designs and multiple methods cannot control for all possible factors that can produce observed results. This means that there can be many feasible and perhaps likely alternative explanations for the observed results of a study. Some designs result
in more alternative possibilities than others, and different designs allow us to rule out different alternatives. Ultimately, if enough different types of studies can be conducted to control for all plausible alternatives (or at least those we can think of), confidence can be placed in the conclusions drawn about the likely explanation for the phenomenon of interest. To accomplish this there must be a series of studies conducted utilizing different designs and methods that taken together can rule out various alternative explanations.

Correlations among observed variables are difficult to interpret in terms of underlying constructs because they can be subject to many influences other than the traits of interest. If two variables share sources of method variance, correlations between them can be caused by the traits of interest or by method variance, or both. For example, it has been hypothesized that the personality trait of negative affectivity (NA) accounts for at least some of the variance in self-reports of many organizational variables (e.g., Brief et al., 1988; Watson, Pennebaker and Folger, 1986). As an unintended influence on questionnaire responses, NA would be a source of method variance for self-reports. Correlations among questionnaire items or scales might be caused, at least in part, by NA rather than intended traits. Several studies have been conducted in which NA was assessed in addition to other variables of interest. This allowed for the partialling of the effects of NA on correlations between other variables, such as job conditions and job satisfaction. Although the magnitude of the NA partialling effect has been a topic of controversy (see Brief et al., 1988; Chen and Spector, 1991), it would seem to be a possible partial cause of correlations among at least some variables of interest in the OB domain.

Unfortunately, the partialling procedure has been unable to indicate what the role of NA might be. It would be premature to conclude that self-reports of job conditions and job satisfaction are caused by NA as there are several possible explanations, some of which suggest that the underlying intended traits are in fact related. For example, Spector, Jex and Chen (1993) found that NA was related to objective measures of job conditions, and several studies have found that NA correlates with job satisfaction (e.g., Brief et al., 1988). One possibility is that NA affects the complexity of jobs chosen and that job complexity leads to satisfaction. This would result in correlations between NA and the other two variables. The conclusion that job complexity was a cause of satisfaction might be correct, despite the fact that both are correlated with NA.

Using additional methods or sources of data can help control some sources of method variance, but others can still remain. For example, suppose one hypothesizes that workload is the cause of physical health problems. Perhaps the easiest way to test this hypothesis would be with a cross-sectional questionnaire study. Employees could be asked to report about their workloads and their physical health symptoms. There would be many explanations for the finding that the two variables were significantly correlated. NA is a possible explanation for why workload and symptoms would correlate. Individuals who are high in NA are likely to perceive heavier workloads and more severe health symptoms than individuals who are low, regardless of the objective reality. This possibility is illustrated with a path-type diagram in Figure 1a.

What is gained by using a different method, such as a physiological measure of health rather than self-report? Even though two self-report measures are not now being correlated with one another, the situation is not much better for concluding that workload affects health. It is still possible that NA is the cause of both variables. Perhaps people who are high in NA have different physiological levels than people who are low. The correlation between self-reports on workload and physiological measures might still be spuriously caused by NA (see the path diagram in Figure 1b).

Can the problem be solved by eliminating self-reports entirely? An objective measure of
workload can be used and correlated with the physiological measure. There is still inadequate certainty if NA is the cause of the correlation or not. As noted earlier, Spector et al. (1993) found that NA is correlated with objective characteristics of jobs. Here again, perhaps high NA people take jobs with greater workload than people who are low in NA. NA could spuriously cause the correlation between objective workload and the physiological measure of health (see the path diagram in Figure 1c).

The situation is not improved much even if the self-report measures are valid indicators of their intended constructs. As shown in Figure 1d, if objective workload leads to perceived workload and physiology leads to perceived symptoms, the correlation between self-reported workload and self-reported symptoms can be spuriously caused by NA.

The problem here for causal inference is not with the nature of the measure but with the nature of the design. A cross-sectional study cannot provide much certainty about the causal connections among variables. To determine cause and effect, a design is needed that assesses variables over time. The strongest design is an experiment in which the independent variable is assessed or manipulated before the dependent variable is measured. When true experiments are not feasible, quasi-experiments or longitudinal observational studies are better able to address causal research questions than a cross-sectional design. With workload, such a design can be found in a study reported by Frankenhaeuser and Johansson (1986). In this study a sample of employees was assessed for catecholamine levels daily as their work hours increased and decreased over a period of several weeks. Catecholamine levels reliably followed length of work schedule, leading to some confidence that work hours had physiological effects. Of course even here there is little certainty about why the effect occurred. It may have been due to a variety
of mechanisms, such as increases in sleep problems or smoking that were caused by increases in work hours. In other words, there might be additional mediational mechanisms that the study did not uncover. Additional studies could certainly test hypotheses about other possible explanations by assessing these variables.

Raggatt (1991) conducted a cross-sectional questionnaire study which found that long work shifts were associated with sleep problems and the use of stimulants. Perhaps these or other factors can explain why long work shifts affect physiology. Despite its methodology, Raggatt’s study provides some interesting results that help enhance our understanding of how long work shifts affect people. They should not be discounted merely because they were from a survey. Taken together the Frankenhaeuser and Johansson (1986) and Raggatt (1991) studies with very different methodologies begin to unravel the complexities of how long work shifts affect people. Additional studies, some of which might use questionnaire data, will be needed to further the field’s understanding of this important OB phenomenon. For example, a longitudinal study might be conducted to see if people switching to longer shifts experience more sleep problems and use more stimulants, and if these two variables are related to catecholamine levels.

Conclusions

The cross-sectional self-report study is one of the major research methods used in OB research. The method has its weaknesses, which are shared with other methods that are held in higher esteem. There are two major problems with this sort of study. First, the use of the job incumbent as the only source of data leaves many alternative explanations for observed correlations other than that the intended traits are related. This problem might be reduced by using different methods and different sources of data, but it will not be eliminated. The second limitation is that cross-sectional designs do not allow for confident causal conclusions. Even the use of structural equation modeling cannot overcome the severe limitations of having all data collected concurrently. There are too many alternative explanations, including that the direction of causality should be reversed and the supposed cause is in fact the effect. For example, it has been suggested that job satisfaction might be the cause of job condition perceptions, rather than (or in addition to) the effect (James and Tetrick, 1986).

Despite the weaknesses of the cross-sectional self-report methodology, this design can be quite useful in providing a picture of how people feel about and view their jobs. They also tell us about the intercorrelations among various feelings and perceptions. This can provide important insights and can be useful for deriving hypotheses about how people react to jobs. Additional methodologies will be needed to fully test these hypotheses, but cross-sectional questionnaires can provide a relatively easy first step in studying phenomena of interest. In many areas of OB (e.g. job characteristics, job stress, leader behavior, and organizational policies and structure) this first step has been taken and it is time to move on to other methodologies to test more thoroughly some of the hypotheses suggested by self-report studies and by theories.

It is unfortunate that questionnaires have developed a bad reputation among many researchers. In part that reputation has developed because too many authors of articles have attempted to draw conclusions from such studies that are not appropriate for the methodology used. A study that attempts to show a linkage between objective job conditions and another variable will not likely provide important insights if it uses a cross-sectional questionnaire method. Too many studies using such methods have at least implicit research questions that cannot be adequately answered with the method used. Furthermore, in articles it is not always made
clear where conclusions based on the data presented end and where speculation begins. There is nothing wrong with speculation as it may provide hypotheses for future research. The tendency to overgeneralize and overinterpret results, however, is not limited to questionnaire researchers. It is not uncommon for laboratory experiments to be generalized to field settings in ways that should be considered as speculation. Should we be suspicious of all laboratory studies because their generalizability to the field is not usually apparent? Laboratory studies can be valuable because they provide relatively easy first tests of hypotheses before they are tried under the more difficult and uncontrolled conditions of the field.

There is much to be learned about work using questionnaire methods. Self-report studies should not be automatically dismissed as being an inferior methodology to others that might have been applied. Where appropriate, their use should be encouraged. The cross-sectional self-report method has provided interesting and valuable data concerning many OB questions in the past, and it will undoubtedly continue to make a valuable contribution to knowledge in the future. However, the methodology used should match the research question asked, and for many OB questions, the cross-sectional self-report study will not provide adequate answers. Unfortunately, the self-report questionnaire is too often asked to answer questions that would be better addressed with other methods. Questions concerning the impact of the job environment (e.g. job characteristics, job stress, leader behavior, organizational policies and structure) cannot be adequately answered with a cross-sectional self-report study.

References


