

Self-report measures:

**An overview of concerns and limitations of questionnaire use in
occupational stress research**

Tiffani Razavi
School of Management
University of Southampton

ABSTRACT

As is the case with many domains of organisational behaviour, occupational stress research has for many years been characterised by the use of self-report methodologies, in particular the written questionnaire, as the primary means of data collection. Reliance on self-report for the measurement of both dependent and independent variables raises concern about the validity of causal conclusions for a range of reasons, including systematic response distortions, method variance and monomethod bias, and the psychometric properties (reliability and validity) of questionnaire scales. Beyond the immediate features of questionnaires, there is also concern about the context in which self-report measures are used, in terms of the design of studies, as well as the statistical treatment of questionnaire data at the analysis stage. Each of these themes is addressed from the perspective of the occupational stress literature, with a view to highlighting key areas for consideration in relation to planning and interpreting research. It is also suggested that most of these issues have wider relevance in the field of organisational behaviour, namely in those domains which focus on perceptual and attitudinal variables, which could benefit from similar methodological attention.

**Self-report measures:
An overview of concerns and limitations of questionnaire use in
occupational stress research**

INTRODUCTION

For over a decade, occupational stress research has been criticised for its reliance on self-report methodologies, in particular written questionnaires, to collect information about both dependent and independent variables (e.g. Kasl, 1987; Kessler, 1987; Frese & Zapf, 1988; Ganster & Schaubroeck, 1991). Participants may be asked to assess and report personal characteristics (e.g. personality), environmental characteristics (e.g. workload, control over the work situation), as well as their own affective responses to that environment (e.g. job satisfaction), mental and physical health.

While the fundamental goal of occupational stress research is aetiological (Kessler, 1987), with the focus on individual well-being, the specific objectives of studies obviously vary. In general terms, however, stress is conceptualised as a complex process, subject to a wide range of potential modifiers (Kaplan, 1996). Studies aim to establish how aspects of the person and/or the environment influence the affective state, physical health or other behaviour of the individual, and therefore necessitate the separate measurement of any component deemed to be theoretically relevant to this causal relationship. Most occupational stress research makes use of self-report measures obtained through questionnaires to assess some or all of the variables of interest. The continuing prevalence of self-report methods demands consideration of the debates concerning their use.

At the most basic level, there is concern about the construct validity of self-report measures. Both theory and research indicate that self-report responses are a product of psychological, sociological, linguistic, experiential and contextual variables, which may have little to do with the construct of interest (e.g. Harrison, McLaughlin & Coalter, 1996; Lanyon & Goodstein, 1997). Because of influences other than item content, it has been pointed out that it is never clear precisely *what* is being measured (e.g. Crowne & Marlowe, 1964; Kessler, 1987; Paulhaus, 1991).

Few would dispute that some constructs are by definition perceptual in nature and therefore are appropriately measured by self-report (e.g. Schmitt, 1994; Spector, 1994), as is the case for values, attitudes and affective responses to the work environment. Howard (1994) asserts that self-report is generally a suitable methodology for the study of human characteristics, and may even be superior to other approaches. Indeed, as qualitative methods gain popularity, it seems likely that the nature of the criticism of self-report methodology in general, and questionnaires in particular, may also shift in focus (e.g. Oakland and Ostell, 1996). Among qualitative researchers, the interview is often preferred above the questionnaire for its potential to provide rich ideographic data, the characteristics of which are in keeping with the interpretive framework. Furthermore, the benefits of the interview as a flexible and probing means of data collection are seen to outweigh the potential advantages of standardised items administered to large sample sizes.

The present review, however, is targeted at the use of self-report data within the broadly positivist framework that is typical of the majority of mainstream occupational stress research. The emphasis of much of the criticism from this

perspective is essentially psychometric: critics of organisational research methodology note that the danger of self-report assessment lies in its misapplication and in the failure to acknowledge the effects of method variance, especially in the measurement and representation of the objective environment. Without due attention to the limitations of the self-report method, or compensation in the form of stronger designs, conclusions may well be open to question (e.g. Kasl, 1987; Spector, 1994).

An understanding of the systematic biases associated with self-report measures is therefore a necessary element of any methodological critique of survey-based research, and of particular relevance when self-report is used to assess *all* variables in a study, a feature not uncommon in occupational stress research. Under such conditions, there is potential response distortion not only of each measure, but also of the correlations between them (monomethod bias or method effect; see Frese & Zapf, 1988; Spector & Brannick, 1995).

This review seeks to address the key concerns and limitations associated with the use of surveys in occupational stress research, in particular the reliance on self-report via questionnaire as the sole or primary data source. Having noted above the general concern of qualitative researchers, this paper focuses on the issues from a quantitative perspective, examining the problems of (1) response distortions; (2) method variance; (3) psychometric properties; (4) design and analysis.

1. RESPONSE DISTORTIONS

Response distortions may be broadly grouped into two categories: response styles and response sets (Lanyon & Goodstein, 1997). Response styles imply bias in a particular

direction regardless of the content of the test items, and include acquiescence, extreme and central tendency responding, and negative affectivity bias, each reviewed below. In contrast, response sets are generally related to content and reflect a conscious or unconscious attempt on the part of the respondent to create a certain impression; the most frequently studied - and of particular importance in organisational contexts - is socially desirable responding.

1.1 Acquiescence response style

Acquiescence response style refers to the tendency to respond positively (“true” or “yes”, regardless of the content of the question, and is particularly problematic for attitude survey research (see Paulhaus, 1991). The methodological implication of acquiescence is the balancing of positively- and negatively-keyed items, such that the bias could be detected through inconsistent responses (e.g. Anastasi & Urbina, 1997; Lanyon & Goodstein, 1997). However, Schmitt and Stults (1985) note that when items with mixed positive and negative wording are factor or cluster analysed, there is a danger that negatively-keyed items will define a single factor, due to the careless responding of even 5% of participants. Thus, although the general conclusion reached in the literature remains that attention to the even distribution of positively and negatively worded items is an appropriate means to counteract acquiescence, it is suggested that such balancing should be coupled with special instructions to participants, as well as caution in the processing and interpretation of data.

Furthermore, Paulhaus (1991) comments that balancing the scoring key effectively controls only one type of acquiescence bias, *agreement acquiescence* (the tendency to agree with all types of items), and is not sufficient to counter *acceptance*

acquiescence (the tendency to endorse all statements, even when contradictory, as true of oneself). Controlling the latter requires the inclusion of conceptual opposites, preferably in a forced-choice format, to allow for a clear picture of personal traits or attitudes.

1.2 Extreme and moderacy response styles

Extreme and moderacy response styles operate in measures which require the use of a rating scale and reflect the tendency for subjects to respond consistently using particular sections of the scale. Research suggests that demographic factors influence extreme responding. For example, women give more extreme responses than men (e.g. Newcomb, Huba & Bentler, 1986; Crandall, 1973; Hamilton, 1968). Age, education and culture have also been shown to relate to moderacy versus extreme response bias (e.g. Marin, Gamba & Marin, 1992; Hui & Traindis, 1989; Stening & Everett, 1984), but there is little support for a link with personality dimensions (see Paulhaus, 1991). Extreme responding can be controlled by altering the response format, but eliminating the rating scale may have other implications for the sensitivity and utility of the measure.

1.3 Negative affectivity bias

Negative affectivity has been shown to account for some of the variance in self-report measures of many variables, including measures of job stress and strain, somatic symptoms, health complaints and life events (e.g. Brief, Burke, George, Robinson & Webster 1988; Watson & Pennebaker, 1989; Vassend, 1989; Brett, Brief, Burke, George et al., 1990; Parkes, 1990). Brief et al. (1988) argue that negative affectivity markedly inflates correlations between stressor and outcome variables, and on this

basis call into question the extent to which relationships detected through self-report research are valid. Although their position gains only mixed empirical support in the literature, depending on the nature of the outcome variable (see, for example, Chen & Spector, 1991; Williams & Anderson, 1994), their recommendation to include negative affectivity measures in self-report research and to statistically control for this response bias in analyses of self-report data is now widely accepted (e.g. Parkes, Mendham & von Rabenau, 1994; Moyle, 1999; Wofford, Goodwin & Daly, 1999).

Recently, the debate about the need to take account of negative affectivity in occupational stress surveys has been re-opened. Some researchers posit that in relation to attitudes and health outcomes in particular, negative affectivity should be measured not to control for potential bias, but because it has a substantive role to play in the study of these issues.

1.4 Social desirability bias

Social desirability bias (see Edwards, 1953) refers to the tendency to answer self-report items in such a way as to deliberately or unconsciously represent oneself in a favourable light. Various factors may motivate respondents to provide responses which they believe are more socially desirable than a truthful answer, including the setting in which the research is conducted (see, for example, Parkes, 1980), and the beliefs of respondents about the purpose of the research. Organisational research is particularly prone to deliberate misrepresentation, as participants may feel that the pattern of their responses will impinge on prospects for promotion, pay or job security. Indeed, in the early days of the use of testing in organizations, Whyte (1956) concluded: “When an individual is commanded by an organization to reveal his

innermost feelings, he has a duty to himself to give answers that serve his self-interest rather than that of The Organization” (p.179). Crowne and Marlowe (1964) add that this kind of advice serves as “self-protection from the entrenched values, norms and predilections of The Organization, to which failure to conform might well result in failure to be hired or promoted” (p.11), but also highlight “the pervasive attitudes of *individuals* towards tests, testing and personal disclosure” (p.12).

The ability of respondents to produce fake answers is well-documented (e.g. Furnham & Henderson, 1982). Deliberate misrepresentation may take the form of “faking good” (making a positive presentation of oneself), or making things look worse than they are when it is believed that such a pattern of answers may produce a desirable outcome, for example, reduced workload (see Anastasi & Urbina, 1997). Evidence suggests that deliberate faking is increased under conditions of high face validity (Bornstein, Rossner, Hill & Stepanian, 1994). However, socially desirable responding is not necessarily a deliberate behaviour; it may also reflect an unconscious inclination to create a positive impression, to avoid criticism or gain positive approval (e.g. Crowne & Marlowe, 1964), or may betray self-deceptive tendencies (e.g. Paulhaus, 1984; Paulhaus & Reid, 1991).

Like other response biases, social desirability is problematic because of its potential contaminating influence on the relationships between variables, including masking and spurious associations. The degree to which such contamination occurs is subject to ongoing debate (e.g. Hogan & Nicholson, 1988; Nicholson & Hogan, 1990; Edwards, 1990; Block, 1990), although most researchers make use of social desirability or lie scales (e.g. Marlowe-Crowne Social Desirability Scale, Crowne &

Marlowe, 1960; Lie scale of the Eysenck Personality Inventory; Eysenck & Eysenck, 1964). Such scales determine the extent to which a participant is susceptible to social desirability bias and therefore allow for statistical control of the response bias.

The use of forced choice and ipsative approaches also reduces social desirability bias but may have methodological implications in terms of scale reliability. With reference to multi-scale instruments, Saville and Wilson (1991) and Bartram (1996) report lower reliabilities for ipsative scores when compared to corresponding normative data, with coefficients becoming gradually lower as the number of scales decreased. However, others have shown that ipsative scales generate inflated reliability estimates (e.g. Johnson, Wood & Blinkhorn, 1988; Tenopyr, 1988). While under certain conditions ipsative and normative measures may be treated as equivalent for practical purposes (Bartram, 1996), fully ipsative single-scale measures do violate the assumptions of reliability models and therefore cannot be subjected to the same forms of psychometric analysis.

Based on the use of social desirability scales, a number of researchers have concluded that, although there are weak to moderate correlations between measures of social desirability and attitudinal variables, social desirability does not have a significant impact on the relationship between variables of interest in organisational research (e.g. Ganster, Hennessey & Luthans, 1983; Spector, 1987, Moorman & Podsakoff, 1992). Nonetheless, as in the case of negative affectivity, there are areas of research which may benefit from the measurement of social desirability as a substantive contributor to outcomes, e.g. individual and organisational values (e.g. Razavi, 2000).

2. METHOD VARIANCE

In general, the above outline of the possible influences of response distortions associated with self-reports suggests that biases are not associated with method alone, but rather arise from the interaction of the nature of the construct, the approach to its assessment and the characteristics of the questionnaire respondent. This conclusion is in keeping with Spector and Brannick (1995), who suggest that method variance is not simply a inherent in a particular method, but that it constitutes systematic effects of measuring a certain characteristic in a certain way. They argue that the variance of questionnaire scores is made up of three components - trait, method, and error variance - and that researchers should use these as a framework for the interpretation of the meaning of a given measure (see also Spector, 1994), employing a conceptually guided multitrait-multimethod approach (Campbell & Fiske, 1959) where possible. Within such boundaries, and in the context of adequately designed research, self-reports can provide meaningful information, and offer a practical, cost-effective means of data collection, particularly suited to the study of perceptual and attitudinal constructs.

As mentioned above, it is in the assessment of environmental characteristics that many researchers advocate the use of methods other than or in addition to self-reports (e.g. Frese & Zapf, 1988; Ganster & Schaubroeck, 1991). Thus, a number of studies make use of reports by others, such as peers, supervisors or observers (e.g. Conway & Huffcutt, 1997; Furnham & Stringfield, 1994; Harris & Schaubroeck, 1988). In general, the multi-rater literature shows varying degrees of agreement, depending on both the construct and the source. Other ratings are susceptible to some of the response distortions of self-reports and may also introduce other biases (see Frese &

Zapf, 1988) but do offer an indication of the components of variance, as suggested by Spector and Brannick (1995).

Many researchers favour the use of objective measures (e.g. Kasl, 1987; Frese & Zapf, 1988; Ganster & Schaubroeck, 1991) as these are free from response bias and, in the case of the work environment, are often a useful basis for the recommendation of interventions. Quasi-experimental designs and naturally occurring experiments (see below), as well as experimental manipulation of environmental conditions, offer the possibility to investigate variations in work characteristics, such as levels of workload, independently of perceptions of those conditions (e.g. Parkes, 1982, 1995). In addition, in terms of outcomes, independent measures of behaviours such as organisational records of turnover and sickness absence can supplement subjective measures and enhance the interpretation of causal relationships within the stress process.

However, despite the attraction of objective measures, they are not exempt from criticism. In contrast to the inflated effects produced by self-reports, objective measures of the work environment may underestimate associations between variables (Frese & Zapf, 1988), as they may lack individual-level variation and, returning again to theory, may fail to capture a conceptually subjective experience or a construct that is essentially perceptual.

3. PSYCHOMETRIC PROPERTIES

The advantages of self-report measures are to a large extent dependent on the psychometric properties of the instruments used in survey research, in particular their

reliability and validity. These properties are fundamental to classic psychometric theory (see, for example, Nunnally, 1978, 1994) and their assessment is critical to the interpretation of results.

3.1 Reliability

Reliability refers to the repeatability or dependability of measurement; hypothetically, changes in levels of a completely reliable instrument would be a reflection of true change in the characteristic of interest. In practice, however, random errors of measurement are never completely absent, and test results may be affected by a range of factors from variations in the format of a test to aspects of the test situation. To the extent that an approach to measurement provides the same results despite these variations, it can be said to be reliable. Notwithstanding debate about the adequacy of single-item measures, the usual recommendation of psychometric theorists is that questionnaire measures should comprise more than a single item (e.g. Nunnally, 1978), as the use of multiple items to assess the construct of interest acts to reduce measurement error in the scale as a whole and thereby tends to increase reliability.

The internal consistency of multi-item scales can be assessed by various split-half procedures, and is most frequently indexed by coefficient alpha, the mean of all split-half coefficients (Cronbach, 1951). While determination of a satisfactory level of consistency is dependent on how a measure is being used, and particularly on the extent to which test scores are used to differentiate between people, a minimum of 0.7 is recommended in the literature (e.g. Nunnally, 1978; Cox & Ferguson, 1994).

In addition to internal consistency, external reliability should also be considered. The stability of a measure across time is most commonly assessed through the mean correlation of scores obtained from repeated administrations of a test (test-retest reliability). There are difficulties associated with test-retest reliability, as a range of factors other than characteristics of the scale (including the interval between administrations, varied test conditions, and practice effects, as well as the hypothetical stability of the construct) may affect the resulting coefficient. Consequently, Nunnally (1978) notes that the test-retest coefficient should not be used to justify a low alpha coefficient.

3.2 Validity

Validity refers to the extent to which an instrument measures what it purports to measure, and not something else. Unlike reliability, it cannot be captured in a single coefficient as it is a multifaceted concept determined by relations with other variables (e.g. Silva, 1993; Anastasi & Urbina, 1997). Although traditionally there are several types of validity (e.g. face validity, content validity, construct validity, predictive validity, concurrent validity) based on the relationship between the test and a specified variable or set of variables and which are still reviewed in psychometric texts (see for example, Rust & Golombok, 1989), recent developments reflect an integrative approach which emphasises validity as a unitary concept. Anastasi & Urbina (1997) argue that construct validity (the degree to which an instrument captures the hypothetical qualities or traits it was designed to measure) essentially encompasses all forms of validity, a view supported by many others, who point out that validation is synonymous with hypothesis testing (e.g. Landy, 1986; Messick, 1989; Silva, 1993). Ultimately, determination of the validity of a measure relies on

evaluative judgement of the adequacy of inferences made from test scores, in terms of both empirical evidence and theoretical rationale (Messick, 1989; Silva, 1993).

3.3 Relationship between reliability and validity

While it is important for instruments to be both reliable and valid, an implication of psychometric theories of measurement error and generalizability (e.g. Cronbach, Gleser, Nanda & Rajaratnan, 1972) is that validity takes priority. Clearly, a valid measure of a construct which is not theoretically stable across time will have an influence external reliability, and low test-retest reliability in itself may be informative. Similarly, not all constructs are theoretically expected to possess perfect homogeneity (Silva, 1993). In terms of meaningful research, then, establishment of the validity of measures is of primary importance, as it represents theoretical viability, but ideally should be supported by recommended levels of consistency and appropriate external reliability.

4. DESIGN AND ANALYSIS

While attention to the psychometric properties of the scales used in questionnaire-based studies is fundamental to good research practice, as pointed out earlier, measurement issues cannot be adequately evaluated in isolation from research design. The primary interest in much of organisational behaviour, and certainly in the field of occupational stress, is in determining causal relationships; the measurement issues discussed above generate concern because they reduce the confidence with which conclusions about causality can be drawn, thereby also diminishing the theoretical and practical value of research findings.

It is now generally accepted that longitudinal designs significantly enhance the interpretation of causality, and, specifically, that they guard against the problems of reverse causation and the influence of third variables, such as negative affectivity (e.g. Kessler, 1987; Schaubroeck & Kuehn, 1992; Spector, 1994).

Nevertheless, much of occupational stress research remains cross-sectional. A recent review (Zapf, Dormann & Frese, 1996) reports that more than 90% of stress research is cross-sectional, and despite increased prevalence of longitudinal studies in the occupational stress literature, there are still relatively few. Furthermore, the review highlights the need for longitudinal research to address the issue of appropriate time lags from both theoretical and empirical perspectives (see also Kessler, 1987), as the timing of measurements has direct bearing on the detection of causal relationships. In addition, measurement of all variables by the same assessment method at each time point is necessary to allow for full exploration of reciprocal and reverse causal effects.

Although longitudinal studies offer clear and important advantages over cross-sectional approaches, there remains a need for such research to incorporate suitable methods of data analysis (see Zapf, Dormann & Frese, 1996). The stress process is multi-faceted and the number of variables potentially influencing stress and strain responses both directly and indirectly as moderators and mediators is large. Multivariate techniques, with their implications for data screening, are therefore essential for the meaningful analysis of data. Hierarchical multiple regression and structural equation modelling are both prevalent in the literature, and appropriate for occupational stress research, subject to the limitations of sample size and distribution.

SUMMARY AND CONCLUSION

Occupational stress research tends to integrate multiple concepts and constructs often of a perceptual or attitudinal nature, in an attempt to establish causal and contingency relationships. In so doing, researchers must necessarily consider of a wide range of methodological issues, beginning with the fundamental question of construct operationalisation and measurement, which more often than not is achieved through survey instruments. The concerns of occupational stress researchers, reflected in the inter-related issues of measurement, design and analysis, must serve as guiding principles for any study attempting to examine relationships between potential stressors and strain responses; indeed this conclusion could well be extended to cover most other domains of organisational behaviour research striving to establish quantitative relationships between predictor and outcome variables. As Schmitt (1994) concludes, "...each researcher must clearly delineate the nature of substantive constructs, potential methods factors, and plan for the measurement and/or control of both in the design and conduct of their research" (p. 397).

Traditional criticisms of self-report methodologies, especially response biases, must be therefore taken into account in the construction of questionnaires, and the analysis and interpretation of data derived from them, but with express reference to the purpose of the study and the nature of the constructs of interest. The choice between self-report and alternative methods of data collection cannot be made in isolation from theoretical considerations. With the ongoing focus on perceptions as theoretically relevant concepts in stress research, and in organisational behaviour more widely, self-report continues to offer both practical and conceptual advantages to organisational researchers which may be enhanced, but not replaced by other

methods. It is therefore not surprising that self-report remains the most common tool, not only in organisational research but in the social and behavioural sciences in general (Harrison, McLaughlin & Coalter, 1996).

REFERENCES

- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). London: Prentice-Hall International.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, 69, 25-39.
- Block, J. (1990). More remarks on social desirability. *American Psychologist*, 45, 1076-1077.
- Bornstein, R.F., Rossner, S.C., Hill, E.L. & Stepanian, M.L. (1994). Face validity and fakeability in objective and projective measures of dependency. *Journal of Personality Assessment*, 63, 363-386.
- Brett, J.F., Brief, A.P., Burke, M.J., George, J.M. et al. (1990). Negative affectivity and the reporting of stressful life events. *Health Psychology*, 9, 57-68.
- Brief, A.P., Burke, M.J. George, J.M., Robinson, B.S. & Webster, J. (1988). Should negative affectivity remain an unmeasured variable in the study of job stress? *Journal of Applied Psychology*, 73, 193-198.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Chen, P.Y. & Spector, P.E. (1991). Negative affectivity as the underlying cause of correlations between stressors and strains. *Journal of Applied Psychology*, 76, 398-407.
- Conway, J.M. & Huffcutt, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10, 331-360.
- Cox, T. & Ferguson, E. (1994). Measurement of the subjective work environment. *Work and Stress*, 8, 98-109.
- Crandall, J.E. (1973). Sex differences in extreme response style: Differences in frequency of use of extreme positive and negative ratings. *Journal of Social Psychology*, 89, 281-293.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnan, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Crowne, D.P. & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.

- Edwards, A.L. (1953). The relationship between the judged desirability of a trait and the probability that it will be endorsed. *Journal of Applied Psychology*, *37*, 90-93.
- Edwards, A.L. (1990). Construct validity and social desirability. *American Psychologist*, *45*, 287-289.
- Eysenck, H.J. & Eysenck, S.B.G. (1964). *Manual of the Eysenck Personality Inventory*. London: Hodder & Stoughton.
- Frese, M. & Zapf, D. (1988). Methodological issues in the study of work stress: Objective vs subjective measurement of work stress and the question of longitudinal studies. In Cooper, C.L. & Payne, R. (Eds.) *Causes, coping and consequences of stress at work*. Chichester: John Wiley & Sons.
- Furnham, A. & Henderson, M. (1982). The good, the bad and the mad: response bias in self-report measures. *Personality and Individual Differences*, *4*, 519-525.
- Furnham, A. & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology*, *67*, 57-67.
- Ganster, D.C., Hennessey, H.W. & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, *26*, 321-331.
- Ganster, D.C. & Schaubroeck, J. (1991). Work stress and employee health. *Journal of Management*, *17*, 235-271.
- Hamilton, D.L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192-203.
- Harris, M.M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43-62.
- Harrison, D.A., McLaughlin, M.E. & Coalter, T.M. (1996). Context, cognition and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes*, *68*, 246-261.
- Hogan, R. & Nicholson, R.A. (1988). The meaning of personality test scores. *American Psychologist*, *43*, 621-626.
- Howard, G.S. (1994). Why do people say nasty things about self-reports? *Journal of Organizational Behavior*, *15*, 399-404.
- Hui, C. & Traindis, H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross Cultural Psychology*, *20*, 296-309.
- Johnson, C.E., Wood, R. & Blinkhorn, S.F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, *61*, 153-162.
- Kaplan, H.B. (1996). Themes, lacunae, and directions in research on psychosocial stress. In Kaplan, H.B. (Ed.) *Psychosocial Stress: Perspectives on structure, theory, life-course, and methods*. San Diego, CA: Academic Press.

- Kasl, S.V. (1987). Methodologies in stress and health: Past difficulties, present dilemmas, future directions. In Kasl, S.V. & Cooper, C.L. (Eds.) *Stress and health: Issues in research methodology*. Chichester: John Wiley & Sons.
- Kessler, R.C. (1987). The interplay of research design strategies and data analysis procedures in evaluating the effects of stress on health. In Kasl, S.V. and Cooper, C.L. (Eds.) *Stress and health: Issues in research methodology*. Chichester: John Wiley & Sons.
- Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, *41*, 1183-1192.
- Lanyon, R.I. & Goodstein, L.D. (1997). *Personality assessment* (3rd ed.). New York: Wiley.
- Marin, G., Gamba, R.J. & Marin, B.V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross Cultural Psychology*, *23*, 498-509.
- Messick, S. (1989). Validity. In Linn, R.L. (Ed.), *Educational measurement* (3rd ed.). New York: American Council of Education, Macmillan & Collier Macmillan.
- Moorman, R.H. & Podsakoff, P.M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, *65*, 131-149.
- Moyle, P. (1999). The effects of transition stress: A relocation study. *Journal of Organizational Behavior*, *20*, 625-646.
- Newcomb, M.D., Huba, G.J. & Bentler, P.M. (1986). Desirability of various life change events among adolescents: Effects of exposure, sex, age, and ethnicity. *Journal of Research in Personality*, *20*, 207-227.
- Nicholson, R.A. & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist*, *45*, 290-292.
- Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J.C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oakland, S. & Ostell, A. (1996). Measuring coping: A review and critique. *Human Relations*, *49*, 133-155.
- Parkes, K.R. (1980). Social desirability, defensiveness and self-report psychiatric inventory scores. *Psychological Medicine*, *10*, 735-742.
- Parkes, K.R. (1982). Occupational stress among student nurses: A natural experiment. *Journal of Applied Psychology*, *67*, 784-796.

- Parkes, K.R. (1990). Coping, negative affectivity, and the work environment: Additive and interactive predictors of mental health. *Journal of Applied Psychology*, 75, 399-409.
- Parkes, K.R. (1995). The effects of objective workload on cognitive performance in a field setting: A two-period cross-over trial. *Applied Cognitive Psychology*, 9, 153-171.
- Parkes, K., Mendham, C. & von Rabenau, C. (1994). Social support and the demand-discretion model of job stress: Tests of additive and interactive effects in two samples. *Journal of Vocational Behavior*, 44, 91-113.
- Paulhaus, D.L. (1984). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- Paulhaus, D.L. (1991). Measurement and control of response bias. In Robinson, J.P., Shaver, P.R. & Wrightsman, L.S., *Measures of personality and social psychological attitudes: Volume 1 of measures of social psychological attitudes*. San Diego, CA: Academic Press.
- Paulhaus, D.L. & Reid, D.B. (1991). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Razavi, T. (2000). Fitting in and feeling fine: The impact of personality on person-organization value congruence and affective outcomes. *Discussion Papers in Accounting and Management Sciences*, School of Management, University of Southampton.
- Saville, P. & Wilson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64, 219-238.
- Schaubroeck, J. & Kuehn, K. (1992). Research design in industrial and organizational psychology. In Cooper, C.L. & Robertson, I.T. (Eds.) *International review of industrial and organizational psychology 1992 Volume 7*. Chichester: John Wiley & Sons Ltd.
- Schmitt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior*, 15, 393-398.
- Schmitt, N. & Stults, D.M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9, 367-373.
- Silva, F. (1993). *Psychometric foundations and behavioural assessment*. Newbury Park, CA: Sage Publications.
- Spector, P.E. (1987). Method variance as an artifact in self-report affect and perceptions at work: Myth or significant problem. *Journal of Applied Psychology*, 72, 438-443.
- Spector, P.E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, 15, 385-392.

- Spector, P.E. & Brannick, M.T. (1995). The nature and effects of method variance in organizational research. In Cooper, C.L, & Robertson, I.T. (Eds.), *International Review of Industrial and Organizational Psychology, Volume 10*. New York: John Wiley & Sons.
- Stening, B.W. & Everett, J.E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology, 122*, 151-156.
- Tenopyr, M.L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology, 73*, 749-751.
- Vassend, O. (1989). Dimensions of negative affectivity, self-reported somatic symptoms, and health-related behaviors. *Social Science and Medicine, 28*, 29-36.
- Watson, D. & Pennebaker, J.W. (1989). Health complaints, stress and distress: Exploring the central role of negative affectivity. *Psychological Review, 96*, 234-254.
- Whyte, W.H. (1956). *The Organization Man*. New York: Simon & Schuster.
- Williams, L.J. & Anderson, S.E. (1994). An alternative approach to method effects by using latent variable models: Applications to organizational behavior research. *Journal of Applied Psychology, 79*, 323-334.
- Wofford, J.C., Goodwin, V.L., Daly, P.S. (1999). Cognitive-affective stress propensity: A field study. *Journal of Organizational Behavior, 20*, 687-708.
- Zapf, D., Dormann, C. & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology, 1*, 145-169.