A critical review of psychosocial hazard measures

Prepared by
The Institute for Employment Studies
for the Health and Safety Executive

CONTRACT RESEARCH REPORT
356/2001

HSE
Health & Safety Executive

A critical review of psychosocial hazard measures

Jo Rick, Rob B Briner, Kevin Daniels, Sarah Perryman and Andrew Guppy
The Institute for Employment Studies
Mantell Building
University of Sussex
Brighton
BN1 9RF
United Kingdom

Health and safety legislation requires that employers regularly conduct risk assessments to identify what in their workplace is a potential hazard to (ie could harm) employee health.

The idea of risk assessment for physical hazards is well-established. More recently attention has focused on the assessment of risk from psychosocial hazards and in doing so measures have been developed or adopted from research to assess the prevalence of workplace stressors.

Whilst much research has been done on stress, there exists no systematic overview of the different types of stressor measures available in the UK, nor is there any consistently recorded information about their relative merits.

This report seeks to fill that gap by identifying a wide range of commonly used measures, assessing the research evidence available on them and providing an overview of their relative strengths.

Conclusions are drawn about the state of knowledge in this area and issues for practice and research.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

HSE BOOKS

First published 2001

ISBN 0 7176 2064 6

The Project Team

A large team of contributors were involved at different stages of this report. Details of the full project team and their affiliations are as follows:

Rob Briner, Senior Lecturer in Organizational Psychology, Birkbeck College, University of London.

Polly Carroll, Occupational Psychology Branch, The Employment Service.

Kevin Daniels, formerly Lecturer in Organizational Behaviour, Sheffield University Management School, now Reader in Occupational Psychology, University of Nottingham.

Andrew Guppy, Professor of Applied Psychology, Middlesex University.

Chris Kelly, Principal Psychologist, Work Psychology Section, Health and Safety Laboratories.

Sarah Perryman, Research Officer, the Institute for Employment Studies.

Jo Rick, Principal Research Fellow, the Institute for Employment Studies.

Claire Tyers, Research Fellow, the Institute for Employment Studies.

The Institute for Employment Studies

The Institute for Employment Studies is an independent, apolitical, international centre of research and consultancy in human resource issues. It works closely with employers in the manufacturing, service and public sectors, government departments, agencies, professional and employee bodies, and foundations. For over 30 years the Institute has been a focus of knowledge and practical experience in employment and training policy, the operation of labour markets and human resource planning and development. IES is a not-for-profit organisation which has a multidisciplinary staff of over 50. IES expertise is available to all organisations through research, consultancy, publications and the Internet.

IES aims to help bring about sustainable improvements in employment policy and human resource management. IES achieves this by increasing the understanding and improving the practice of key decision makers in policy bodies and employing organisations.

Acknowledgements

Our thanks go to the many people who helped in the preparation of this report. In particular to Clare Duffy and Sue Waller our project managers at HSE for there continued support and guidance.

Thanks also go to Caroline Beaumont and Becky Lincoln for their work in preparing the final draft of this document.

Contents

1. Introduction

There is a rapidly growing body of research on the management and control of workplace stress. Some of it has attempted to categorise the types of stress management initiatives organisations can undertake. These categorisations often include the ideas of prevention at source, providing individuals with skills which may help them deal with stress problems, or treating those who have been harmed (see, for example, Ivancevich, Matteson, Freedman and Phillips, 1990). One approach which has gained in popularity and which may help organisations determine the kind of stress intervention to undertake has been to try and measure workplace stressors through the use of self report questionnaires.

In effect, this review is concerned with the measurement of workplace stressors. **Throughout the document the term 'psychosocial hazards' is used to refer to work characteristics which could equally be termed 'stressors' or 'sources of stress'.** Why then use 'psychosocial hazards'? The rationale has three parts:

- 'Stress' is generally acknowledged to be a broad umbrella term for a wide range of different experiences and conditions. It is generally accepted to be a vague concept consistently beset by problems of definition. In contrast the focus of this study is very clearly that of the psychosocial aspects of work that have the potential to adversely impact an individual's mental and physical well-being.
- Stress tends to bring with it a plethora of sometimes unhelpful ideas and expectations about stress management, whereas the legislation which governs psychological health at work and the requirements it places on employers is concerned specifically with risk assessment and the monitoring and control of hazards
- The HSE has done much to clearly establish a risk assessment framework in relation to physical hazards. This framework when applied to psychosocial hazards helps to make explicit the steps which may also be involved in assessing workplace risk. See, for example, Cox, Leather and Cox (2000).

In May 2000, the Health and Safety Executive (HSE) commissioned this study to look at what is known about the

quality of psychosocial hazard measures. This chapter looks first at the objectives for the research and then at the broader research context for the study.

1.1 Research objectives

The purpose of the research is to provide a critical review of current psychosocial hazard measures. The research was funded by the HSE because whilst there is a large and fast growing literature on stress, there is no systematic overview of the numerous measures of psychosocial hazards, nor is there consistently recorded information about their relative merits: Hence the need to take stock of the activity in this area and draw conclusions about its relevance to employers and others who may wish to use such measures.

To do this, the research needed to fulfil three tasks. First it needed to identify the methods or measures currently used to assess psychosocial hazards in the workplace.

Second, having identified existing approaches to psychosocial hazard measurement, the major part of the research needed to provide a comprehensive review of each of the measures identified. This should include where appropriate:

- evidence of reliability (*ie* the consistency of measurement) against recognised standards
- evidence of validity (*ie* the meaningfulness or relevance of measurement) in relation to risk assessment (identification of real hazards) against recognised standards. Where the data exists this should include evidence of: face; content; construct; and predictive validities
- evidence in relation to the validity of the underlying theoretical foundations of measures and how that relates to current thinking.

Last, the utility of different approaches needs to be considered. Utility can refer to the costs and benefits associated with use of a particular tool. In relation to this study, however, utility refers to the 'ease of use' of a measure and therefore needs to address issues such as training requirements, administration and completion time; the interpretation of the results; and, the ease with which findings can be related to specific actions in the workplace.

1.2 The risk management framework

Work in the area of hazard control and risk management provides a framework which seeks to enhance and improve occupational health and safety practice. Cox (1993) has argued that the

principles of regulations such as the Control of Substances Hazardous to Health Regulations (originally made in 1988, but subsequently amended and remade several times) can be effectively employed to manage psychosocial hazards and the psychological harms which may be a consequence of such hazards. The implications are that psychosocial hazards can be managed in much the same way as physical hazards and similar risk assessment procedures can be used in their identification and control in the workplace.

The guidelines for using this framework can be seen as a cycle of activities and involve six steps as follows:

- identification of hazards
- assessment of associated risk
- implementation of appropriate control strategies
- monitoring of effectiveness of control strategies
- reassessment of risk.

Risk assessment is defined by the HSE (1998) as:

> *'nothing more than a careful examination of what, in your work, could cause harm to people, so that you can weigh up whether you have taken enough precautions or should do more to prevent harm. The aim is to make sure that no one gets hurt or becomes ill.'*

In the context of this study, we understand risk assessment to consist of three key elements:

- **Hazard** — anything that has the potential to cause harm.
- **Harm** — the impact of a work hazard.
- **Risk** — the chance that someone will be harmed by a hazard.

Risk assessment should also, according to Cox (1993): 'both offer an explanation of and quantify the hazard-harm relationship', *ie* it should reveal how and why there is a hazard-harm relationship as well as the extent of that relationship.

There are differences between psychosocial and physical hazards (Briner and Rick, 1999; Rick and Briner, 2000) that not only have implications for how psychosocial risk assessments are conducted, but also the procedures we use to judge the accuracy of psychosocial risk assessment measures.

1.3 What do organisations do?

While it is clearly important that organisations conduct risk assessments for psychosocial hazards, what do we know about what organisations actually do in this area? In general, it appears that although psychosocial hazards and harms may be assessed in

some way, this is not necessarily done within a risk management framework. Indeed is appears likely that what organisations do is engage in wide range of rather different kinds of activities which, either deliberately or otherwise, assess and manage psychosocial hazards and harms.

One important aspect of conducting a risk assessment for psychosocial hazards is the extent to which organisations actually complete each of the different elements of the risk assessment process. Previous research by Rick, Young and Guppy (1998) into managing work based trauma concluded that risk assessments for traumatic incidents varied considerably across case study organisations participating in the research. What risk assessment existed tended to focus on numbers of incidents (*ie* stressors/hazards) but failed to assess the consequent harm associated with such incidents.

Similarly, case study research by IES (Rick, Hillage, Honey and Perryman, 1997) into employer responses to stress at work also identified some instances of inferring harm from the existence of possible hazards. In other words, employers *assumed* employees were being harmed because they could measure the existence of certain stressors in the workplace.

It is clear that, in some cases, organisations measure what are presumed to be harms (*eg* using generic mental health measures such as GHQ-12) and, having identified poor mental health/well-being, then infer evidence of workplace hazards. On the other hand, organisations may measure hazards (using measures of workplace stressors) and infer that psychosocial harm is occurring as a consequence. Cox and Griffiths (1996) point out that identifying problematic areas of work alone is not sufficient for the identification of psychosocial risk factors in the assessment of risk; evidence of associated harm is also required. Other organisations may try to build a more complete picture by assessing both hazards and harms and attempting to establish what kinds of relationships, if any, exist between them.

There is also an issue around the extent to which organisations adopt these risk management frameworks to assess psychosocial hazards and harms. Recent research by IRS, for example, found that 74 out of 126 employers responded positively to the question: 'has your organisation tried to identify the causes of work related stress?' However, only 44 (*ie* 34 per cent) had used health and safety risk assessments to help them do this. In other words, organisations may be assessing what are, in effect, hazards and harms but not doing so within a risk management framework.

Also, organisations may in practice manage psychosocial hazards and their impacts but do this through other kinds of activities. Recent research by the Health and Safety Laboratory (Kelly, Sprigg and Sreenivasan, 1998) found that some managers in small

and medium sized employers (SMEs) were doing things about stress, some of which could be viewed as primary interventions. However, it was only as a result of taking part in the focus groups for the research that some SME managers realised such activities counted as doing something about stress (as opposed to simply being good management practice). Likewise, other practices such as flexible working or employee attitude surveys may help with identifying and managing psychosocial hazards and harms but may not be considered in this light by organisations. Indeed, many of the principles of good management in Investors in People (IiP) might otherwise be labelled as primary stress management.

This would suggest that organisations use methods other than the kinds of psychometric scales reviewed in this research to identify workplace stressors. This has implications for the current project. A complete review of each and every method of psychosocial hazard assessment might include, for example, checklists of job or environment characteristics, observations of employees, focus groups, data from attitude surveys, measures of output, *etc*. The use of such techniques is likely to vary across different types of workplaces and different kinds of specific hazards. For example, research into violence at work has identified specific hazards, such as cash handling or certain aspects of working with the public. Checklists have been developed to identify and minimise risks in relation to these hazards and such checklists are widely used in certain settings, *eg* the violence risk assessment checklist promoted by UNISON or 'Violence and Aggression to Staff in Health Services' (HSE, 1997). Attempts to review the whole area would be an enormous task. The specific focus of this review is on those psychometric scales or questionnaires designed to measure workplace stressors.

1.4 What psychosocial hazard measures are available?

To complete this type of review, it is important that the range of psychosocial hazard measures currently available are accurately identified.

Broadly speaking, measures fall into three categories:

- *Generic measures* which can be used across any work setting.
- *Occupation specific measures*, designed for use in particular workplaces, *eg* measures specific to the health service *etc.*
- *Hazard specific measures*, for example, work on violence has led to the production of, for example, checklists which aim to help organisations identify and improve situations or settings that contribute to the risk of workplace violence.

Working in the United States, Quick, Quick, Nelson and Hurrell (1997) undertook a review of objective and diagnostic measures of work stressors. The review included a range of stress related

diagnostic instruments. This type of review is helpful in identifying some of the existing measures. However, not all measures are available in the UK or would cross from US to UK work cultures. A sense of what is available in the UK can be gleaned from looking at publishing catalogues. However, an examination of the catalogues of five major test publishing houses (Thames Valley Test Company; NFER Nelson; SHL: Psychological Corporation; and ASE) revealed only one product, the Occupational Stress Indicator as available specifically for the measurement of workplace stressors. Generic measures of harm were also available (*eg* General Health Questionnaire [GHQ-12]) which measures poor levels of psychological well-being.

Other measures in existence were identified primarily through the literature. Examples include the Pressure Management Indicator, the Job Content Questionnaire and the measures of job characteristics developed for the Department of Health as part of the NHS Workforce Initiative (Haynes, Wall, Bolden, Stride, Rick, 1999).

The purpose of this report is to review the evidence about measures for which reliability and validity data are available. It does not, therefore, cover proprietary instruments, where results have not been published in the literature.

Chapter 2 considers in more detail the ways in which psychosocial hazard measures are used and the extent to which they fulfil the requirements for a risk assessment.

Chapter 3 explores in detail the methodology used to identify current psychosocial hazard measures and the range of measures included in the review.

Results from the review of main measures are presented in Chapter 4 and information on additional measures is in Chapter 5. Prospective conclusions are given in Chapter 6.

2. Effectiveness of Measures of Psychosocial Hazards (Stressors)

2.1 What are measures of psychosocial hazards?

This research evaluates the reliability and validity of a number of measures of psychosocial hazards (stressors). Psychosocial hazards are aspects of the work environment that are thought to have the potential to affect negatively the well-being of employees. The negative effects of psychosocial hazards are often referred to as 'strain'. A great many aspects of the work environment have the potential to cause strain, and hence measures of psychosocial hazards range widely, taking in factors such as job demands, the nature of relationships with co-workers, and the amount of control employees have over work processes.

There are a number of ways in which psychosocial hazards in the work environment could be assessed, including ratings made though observation, measures of production such as output, and through interviews. However, despite extensive searching, it quickly became apparent that the type of psychosocial hazard measures which are formally available, and about which it might be possible to collect reliability and validity data on performance, were almost exclusively self-report measures in which employees are asked to make various kinds of quantitative ratings of particular psychosocial hazards. Self-report measures are by far the most common type of psychosocial hazard measurement. **As a result, it was decided at an early stage of the research to focus on self-report measures of psychosocial hazard.** They are also important as they are based on the widespread assumption within the stress literature that it is employees' **perception** of psychosocial hazards that plays the key role in producing strain. In other words, whether or not any potential psychosocial hazard actually impacts on employee well-being depends to a large extent on the way in which employees perceive that psychosocial hazard. We return to this issue later.

Measures of psychosocial hazards can be quite general in that they attempt to assess the general level of perceived psychosocial hazards in the workplace, or they can be quite specific and focus on perhaps one or two particular types of psychosocial hazard. Whilst all these measures are self-report they differ in terms of the

kinds of items presented and responses required. For example, some items may be very similar to those found in widely-used employee attitude or opinion surveys, where the respondent is required to indicate the extent to which they agree or disagree with a statement such as: 'in this job there is a great deal to do'. Other kinds of items may ask about the stressfulness or otherwise of potential psychosocial hazards more directly though a question such as: 'to what extent do you find each of the following to be stressful?'. Such questions are followed by a list of potential psychosocial hazards such as 'workload' and 'relationships with colleagues' and respondents are required to indicate the extent to which they feel each of the potential psychosocial hazards presented is stressful. Yet others ask about the number of times an individual has experienced a certain situation in a given time frame. There are other types of both items and response scales, and these will be discussed in more detail later. However, it is worth noting that measures differ in the extent to which they attempt to assess the magnitude of a hazard, its prevalence, or both.

2.2 Why do organisations measure psychosocial hazards?

As with many organisational practices the reasons why organisations choose to measure psychosocial hazards are many and varied. In essence, these measures provide information about the way in which individual or groups of employees perceive aspects of their work. Some of the main reasons organisations may collect such information are described briefly below.

Some reasons why organisations collect information about psychosocial hazards (stressors)

- identify potential stress problems
- seek causes for existing problems which may be a consequence of work psychosocial hazards
- examine the possible effects of organisational changes on perceptions of psychosocial hazards
- help focus and target interventions
- identify particular groups who may be experiencing difficulties
- compare psychosocial hazard scores with other organisations and other employees
- provide a baseline to track changes over time
- evaluate the effectiveness of interventions
- give feedback to individuals or groups on their perceptions of psychosocial hazards
- as part of more general employee attitude or opinion surveys
- alert line managers to problems or potential problems
- assess potential hazards as part of a risk assessment.

Organisations may therefore vary widely in their reasons for measuring psychosocial hazards. In addition, an organisation may wish to measure psychosocial hazards to meet multiple goals — from providing individual feedback to targeting interventions.

2.3 How are they used?

There is little systematic information about the ways in which organisations use psychosocial hazard measurements. It is possible, however, to identify a number of likely approaches to their use.

2.3.1 Hazards and well-being

First, measures of psychosocial hazards are likely to be used along with other kinds of measures that assess other aspects of employees' attitudes, feelings, and behaviours. For example, measures of strain (or well-being) may be used alongside measures of psychosocial hazards to determine whether or not there are associations between an employee's, or groups of employees', perceptions of psychosocial hazards and their reports of strain. Organisations may also be interested in the possible links between perceptions of psychosocial hazards and other employee perceptions and behaviours, such as commitment to the organisation, performance, and absence. While measuring psychosocial hazards may be useful to some extent in itself, organisations may also need to know if perceptions of psychosocial hazards are related to employee feelings and attitudes.

2.3.2 Psychosocial hazards and employee opinion surveys

A second feature of the way in which organisations use psychosocial hazard measures concerns the context of the administration of the measures. Psychosocial hazard measures may be included in a much broader employee opinion survey which also measures many other kinds of attitudes. On the other hand a more focused approach may be adopted in which a questionnaire which focuses entirely on psychosocial hazards and their possible effects is administered. The latter type of approach is sometimes described as a 'stress audit'. A further issue related to administration concerns the sample of employees who are asked to complete the questionnaire or survey. Samples could include the whole workforce, a representative sample of the whole workforce, particular groups of employees, or even individual employees.

2.3.3 Different types of psychosocial hazard measure

A third point in considering the use of psychosocial hazard measures is the kinds of choices that are made from the many different kinds of psychosocial hazard and attitude measures available, as will be discussed in more detail later. Some of these are available commercially and others are in the public domain and can be used free of charge. Organisations may also choose to develop their own methods of measuring psychosocial hazards. The approach adopted by organisations may depend on a number of factors including the aims of psychosocial hazard measurement, the availability of internal and external expertise, and other resources.

2.3.4 Using the results from measures of psychosocial hazard

Last, a key issue in using psychosocial hazard measures is what is done with the results. As with organisational attitude surveys, the way in which information is analysed, fed back and acted upon may be the most important factor in their success or otherwise. In the case of psychosocial hazard measurement or stress audits, much may depend on how the survey was developed and with whom, which can vary widely. For example, a stress working party comprising a number of stakeholders such as trades unions, human resource managers, occupational health professionals, and health and safety representatives may have responsibility for the design of the survey, its analysis, and the feedback. In such a context, the feedback is likely to be given to many different stakeholders who may use if for many different purposes. Another approach may involve an initiative from the human resource management or personnel department who design and administer the questionnaire and then decide on the nature of the analysis or feedback.

The analysis itself can be largely descriptive, simply showing how employees have responded to the items. On the other hand, the analysis may relate the measures of psychosocial hazards to other measures, such as strain, or look at the influence of other factors (*eg* tenure, grade) on how psychosocial hazards are reported. Based on the results and on the feedback, organisations are then faced with a number of choices about what, if anything, they should do in response to the results.

2.4 Does measuring psychosocial hazards work or help?

At the present time there is very little available information about whether or not using these kinds of measures is actually effective in helping to reduce or manage stress problems in organisations, or how effective they are compared to other methods of assessing

psychosocial hazards (*eg*. interviews, observation *etc.*). This lack of information is perhaps surprising and disappointing given how widely they appear to be used and the importance of the problems they aim to address. However, as indicated earlier, it is not uncommon for quite widespread organisational practices to have received relatively little evaluation. Given the multiple aims and reasons behind measuring stressors described earlier (Section 1.3), the answer to whether or not measuring psychosocial hazards helps depends to a large extent on why stressor measurement is being undertaken.

So while little is known about whether stressor measures are effective in terms of helping an organisational stress management process, we can address and evaluate some of the more fundamental assumptions behind the use of these measures. As indicated earlier, it is thought that employees' perceptions of psychosocial hazards are key in determining the effects of psychosocial hazards. One assumption therefore underlying the use of these measures is that they are capable of reliably and accurately measuring employee perceptions of psychosocial hazards: for example, do stressor measures actually measure psychosocial hazards consistently or do they pick up on other kinds of perceptions which may not be relevant? A second assumption is that the perception of psychosocial hazards is **causally related** to certain negative outcomes. For example, if an employee reports high workload or that they find their workload to be stressful, will it actually **cause** strain or lower levels of well-being at some point in the future? These are two examples of aspects of reliability and validity when applied to stressor measurement.

Risk assessment is vital for the effective management of stress in organisations. The first step of any risk assessment is the identification of hazards. It is increasingly common for that identification process to use the psychometric measures defined at Section 2.1. However, if hazard identification, and ultimately risk assessment based on this method, are to be successful, then it is essential that the measures used are reliable and valid — this is the central purpose of this report.

3. Methodology

This chapter describes the process developed for identifying measures of psychosocial hazards and the research papers which contain information about the reliability and validity of those measures. In the main, evidence on reliability and validity of different measures was sought from peer reviewed published research.

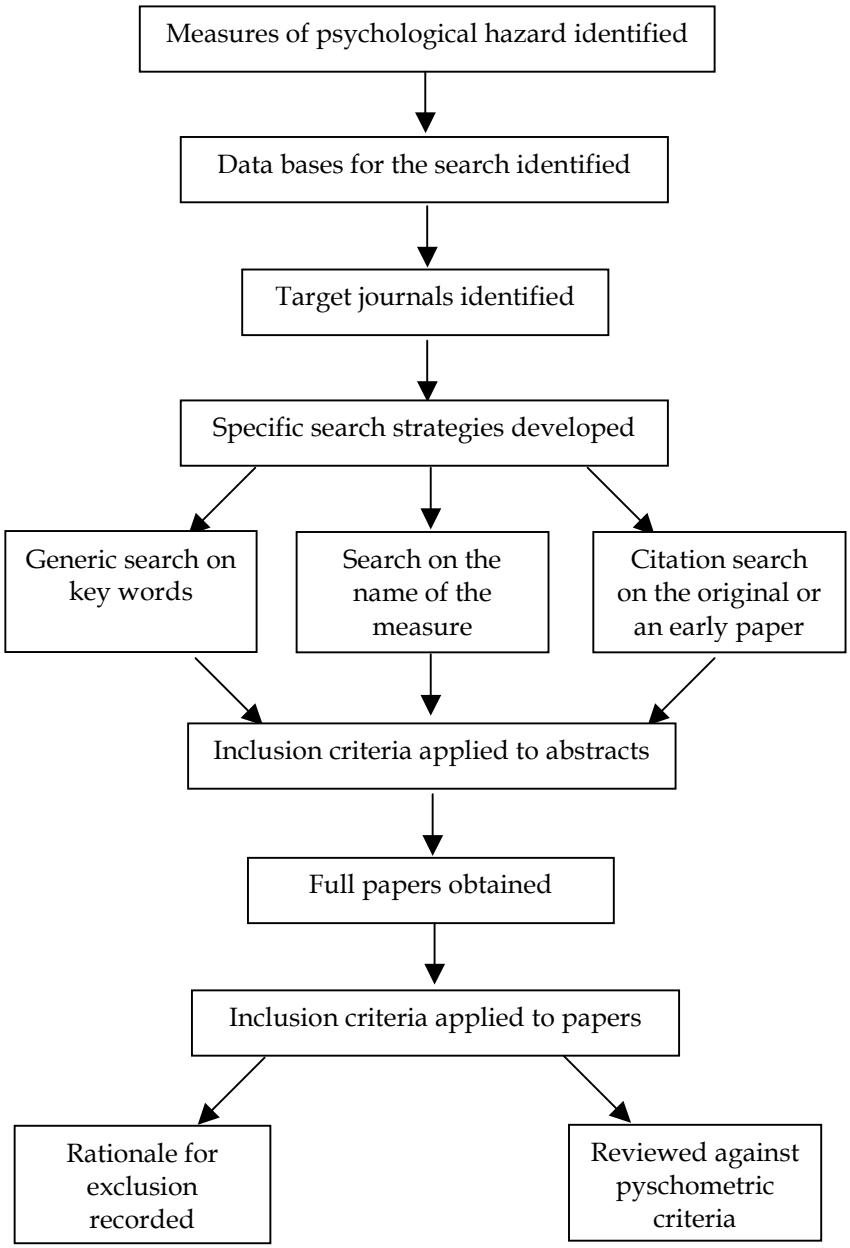
The methodology covers four distinct aspects of the work:

- First, the search procedure for identifying research is described.
- Second, results are given in terms of numbers of papers identified as relevant to the research.
- Third, the process by which research evidence was assessed is described.
- Fourth, the psychometric criteria used are explained.

It should be noted that most of the evidence for the reliability and validity of these measures does not come from research in which a primary aim was to assess their reliability and validity. Rather, most of the available evidence is taken from studies which used the target measure to assess a range of work-related factors.

At an early stage it became evident that identifying evidence from papers was not as straightforward as had initially been anticipated. To undertake a thorough review of the evidence in this area it was recognised that a number of search strategies would be required. These strategies were then assessed at each stage of the process and developed as required to make the search as comprehensive as possible. The review process is represented in Figure 3.1.

Figure 3.1: The review process



Source: IES, 2000

3.1 Identifying measures of psychosocial hazards

3.1.1 The search procedure

The starting point was to identify measures of psychosocial hazards that could be included in the review. This was done first by consultation within the research team and the HSE steering group. An initial meeting of the project team and HSE staff generated a list of measures and approaches. This was then circulated to the project team and steering group for further

consideration and checked against other lists of measures such as that supplied in Quick, Quick, Nelson and Hurrell (1997). The aim at this stage was to be as inclusive as possible, so references to groups of measures were also included for further investigation. The following measures were identified as a starting point:

- Chatman/the Culture Inventory
- Effort-Reward Imbalance (Siegrist)
- Frese
- Hassles and Uplifts Scale
- House *et al*. — Measures of Role Stressors
- Jackson's Measures of Demand and Control
- Job Content Questionnaire
- Job Stress Survey
- Karasek's Measures of Demand and Control
- Life Events Scale
- Michigan Stress Assessment
- NIOSH Generic Job Stress Questionnaire
- Occupational Pressure Inventory
- Occupational Stress Indicator
- Occupational Stress Inventory
- Organisational Stress Health Audit (OSHA)
- Pressure Management Indicator
- Quality of Employment Survey
- Role Experiences Questionnaire
- Stress Audits
- Stress Diagnostic Survey
- Stressors Checklist
- The Job Diagnostic Survey
- The Stress Profile
- Work Environment Scale
- Work Related Strain Inventory.

Having identified this list of measures or areas for further investigation, work was then undertaken to identify original references which could be used in the search.

3.1.2 Identifying data bases for the search

Having identified a wide range of measures, the focus then turned to more specific search strategies. This research used licensed

databases available mainly through academic libraries and used widely in research for the identification of studies on a particular topic. Abstracts (or summaries) from most relevant academic journals are collected on the databases and were identified through a search on key words. The following data bases were identified as most relevant: Psyclit, Medline, and Web of Science (the replacement for BIDS).

Psyclit

Psyclit contains citations and summaries of journal articles, book chapters and book literature in psychology, as well as the psychological aspects of related disciplines, from 1987 to the present, from UK and international periodicals.

Medline

Medline is produced by the US National Library of Medicine and it covers all aspects of medicine, including psychology, from the international academic literature. Since 1982, 3,600 journals have been considered and more than a third of a million records are added each year.

Web of Science

The Web of Science has a Social Sciences Citation Index, which is a data base of the research and scholarly articles in around 1,700 core journals in psychology, business and management, and social policy. It contains over 2.3 million references.

The Web of Science also allows a search to be conducted on the bibliographies or reference lists of the articles indexed in the data bases, so that it is possible to find work which cites particular authors/papers. The citation indices cover the period 1981 to present.

3.1.3 Journals included in the review

In addition to targeting data bases, specific journals were identified to help focus the search, see Table 3.1. These were selected on the basis that they were:

- dominant in relevant disciplines, *eg* occupational health, occupational medicine, applied psychology, *etc.*
- known to publish work in relevant fields
- Peer–reviewed (*ie* papers are blind-reviewed by other researchers and only accepted if of sufficient quality — see Section 4.1.1. for a description).

Table 3.1: Journals selected for inclusion in review

Academy of Management Journal	Journal of Occupational Health Psychology
Applied Ergonomics	Journal of Organizational Behavior
Behavioral Medicine	Journal of Personality and Social Psychology
British Journal of Health Psychology	Journal of Psychosomatic Research
British Medical Journal	Journal of Vocational Behaviour
Ergonomics	Occupational and Environmental Medicine
Human Factors	Occupational Health and Industrial Medicine
Human Relations	Occupational Medicine (Oxford)
International Journal of Industrial Ergonomics	Occupational Medicine (State of The Art Reviews)
International Journal of Industrial Organization	Organization Science
International Journal of Industrial Psychology	Organizational Behavior and Human Decision Processes
International Journal of Stress Management	Personality and Individual Differences
Journal of Applied Psychology	Personnel Psychology
Journal of Applied Social Psychology	Personnel Review
Journal of Behavioral Medicine	Psychosomatic Medicine
Journal of Community and Applied Social Psychology	Social Science and Medicine
Journal of Health and Social Behavior	Stress Medicine
Journal of Management	Work and Occupations
Journal of Occupational and Environmental Medicine	Work and Stress
Journal of Occupational and Organizational Psychology	Work Employment and Society

Source: IES, 2000

It is acknowledged that this approach might miss some articles. However, any other important work would have been picked up through the citation searching that comprised of one of the main search strategies.

3.1.4 Search strategies developed

The search strategy had three stages:

- a general search on likely combinations of key words, *eg* 'psychosocial AND risk assessment'
- a search on the names of the measures
- a citation search on the original paper/manual which first describes the measure or approach.

3.1.5 Inclusion criteria

On the basis of the abstract, a decision was taken about whether the paper was likely to meet the following inclusion criteria:

- *Sample size*: minimum of 100 individuals. This is the smallest sample which can support reasonable multivariate analysis during the development and statistical assessment of the tool.
- *Sample population*: working adults, asked about work. A preference for multi organisation studies over single organisation ones was proposed, as was the preference for UK study groups, However, no papers were excluded for either reason alone at this stage.
- *Sampling methodology*: Full population studies, random or systematic sampling were sought in preference to convenience sampling.

3.2 Targeting appropriate material

Using ten measures chosen at random from the initial list, the number of papers for consideration was recorded from the three data bases. A pilot search was conducted using three different cut off points, *ie* 1995, 1990 and using the full extent of each data base's holdings.

It was estimated that using the full data bases' contents would have generated several thousand abstracts for sifting, many of which would have been irrelevant. As the aim of the project is to identify material on measures and techniques in current use, a cut off date of 1990 was imposed. There was little difference in the estimated number of papers using the 1995 and 1990 cut offs, and the range of material identified. Measures which had been published before 1990 were not necessarily excluded, providing articles had been published about or using them after 1989.

In addition to limiting the study to measures in current use, this also ensured that the review was manageable within the time scale and budget. For similar reasons the review was limited to those articles published in English.

As the interest was in original research data of reasonable quality which provided evidence about the validity and reliability of the measures, the search was limited to articles describing original research, published in a peer-reviewed journal (see list in Table 3.1) or otherwise in the public domain in a peer acceptable form, *eg* a government report. Review articles, letters, news items or commentaries and conference abstracts not subsequently found to be published in a peer-reviewed journal were excluded, as were reports not in the public domain from commercial organisations which produce tests and measures.

3.3 Search and review process

Additional measures were identified in the first, most general search and were added to the list of potential measures and approaches to evaluate.

The second step was a search on the names of the measures *etc.* identified so far, in each of the three data bases. The full abstract for each of the papers was then downloaded. In the version of Medline available, it was not possible to limit the search to the preferred journals, and the journal sift was carried out by hand.

Once the original papers, books or manuals which described the development of the tool and/or its first use were identified, a citation search in the Web of Science was carried out and the results downloaded. As with Medline, the Web of Science citation search feature does not allow a search to be restricted to particular journals, and again, this stage was completed by hand.

Additional material was also sought from a wide range of relevant books which would not have been identified though the data bases. These books were identified through examining the reference lists in papers and through the specialist knowledge of the research team.

The initial searches generated several thousand potentially relevant abstracts. They were all assessed against the inclusion criteria and a much shorter list of full papers was followed up (see Table 3.2).

After reading through the abstracts it became clear that some measures (not just papers) should be excluded on the grounds that the approach was not directly relevant to work, *eg* the Life Events Scale. Those papers which reviewed psychosocial/ psychological factors in general, and not in relation to work in particular, were also excluded at this stage.

3.4 Search results on tool names

Table 3.2 shows the number of papers which mention the tool by full name in the abstract, title or key words.

There was a large degree of overlap between the three different kinds of search, and also between data bases. A large number of duplicates were identified and rationalised.

Table 3.2: Search results on name of measure

Measure	Bids/WoS			Medline			Psyclit			Total no. of hits	No after duplicates identified
	1/1/1990-17/7/2000			1/1/1990-17/7/2000			1/1/1990-18/7/2000				
	HITS	OUT	IN	HITS	OUT	IN	HITS	OUT	IN		
Effort-Reward Imbalance	9	5	4	19	—	5	11	5	6	15	8
Interpersonal Conflict at Work Scale	0	—	—	0	—	—	1	0	1	1	1
Job Content Questionnaire	5	4	1	12	11	1	4	1	3	5	4
Job Environment Scale	0	—	—	0	—	—	0	—	—	—	—
Job Stress Survey	1	0	1	2	0	2	4	0	4	7	4
Measures of Demand and Control*	23	15	8	10	5	3	10	3	7	18	10
Measures of Role Stressors	14	9	5	1	1	0	26	18	8	13	11
NHS Workforce Initiative/scales	0	—	—	10	10	0	10	—	—	1	1
NIOSH Generic Job Stress Questionnaire	0	—	—	1	0	1	0	—	—	1	1
Objective Work Characteristics	0	—	—	1	—	0	0	—	—	—	—
Occupational Stress Indicator	39	18	21	11	11	0	36	20	16	37	26
Occupational Stress Inventory	5	2	3	5	5	0	7	4	3	6	4
Organisational Constraints Scale	0	—	—	1	—	1	1	0	1	2	1
Organisational Stress Health Audit	0	—	—	0	—	—	0	—	—	—	—
Pressure Management Indicator	0	—	—	1	—	1	1	0	1	2	1
Quantitative Workload Inventory	0	—	—	1	—	1	1	0	1	2	1
Role Ambiguity/Conflict Measures	63	37	26	30	30	0	26	17	9	35	31
Stress Audits	4	0	4	27	27	0	2	0	2	6	4
Stress Diagnostic Survey	0	—	—	0	—	—	3	2	1	1	1
The Job Diagnostic Survey	13	4	9	4	4	0	16	10	6	15	13
The Stress Profile	7	6	1	27	27	0	6	5	1	2	1
Whitehall (II) Studies	52	36	14	90	85	5	9	4	5	24	15
Work Environment Scale	5	5	0	17	17	0	6	4	2	2	2

*(Demand and control) and (measure or tool or scale or checklist or survey)

Source: IES, 2000

3.5 Search results on citations

With the downloaded abstracts from the citation searched it was possible to identify and exclude those papers already found through the searches on tool names. In addition, some papers appeared in the citations of more than one tool. The results of the citation searches are presented in Table 3.3.

A full copy of the new papers identified were then obtained for the next stage of the reviewing process.

Table 3.3: Citation search results

Measure	Original Reference(s)	Hits (all journals)
Effort-Reward Imbalance	Siegrist J and Peter R (1994), 'Job stressors and coping characteristics in work related disease. Issues of validity', *Work and Stress,* Vol. 8, 2. pp 130-140	4
	Siegrist J and Peter R (1996), *Measuring effort-reward imbalance at work. Guidelines*, University of Dusseldorf	2
Frese Measures	Frese M (1985), 'Stress at work and psychosomatic complaints — a causal interpretation', *Journal of Applied Psychology,* 70, 2, pp 314-328	60
Interpersonal Conflict at Work Scale	Spector P E and Jex S M (1998), 'Development of four self report measures of job stressors and strain…', *Journal of Occupational Health Psychology,* Vol. 3, 4, pp 356-367	3
Job Content Questionnaire	Karasek R A (1985), *Job content questionnaire and users' guide*, Los Angeles: University of Southern California Department of Industrial and Systems Engineering	48
Job Environment Scale	Caplan R D, (1975), *Job demands and worker health: main effects and occupational differences*, Institute of Social Research, University of Michigan	189
Job Stress Survey	Spielberger C D (1994), *Professional manual for the job stress survey,* Odessa, FL, Psychological Assessment Survey	
Measures of Demand and Control	Karasek R A (1979), 'Job demands, job decision latitudes, and mental strain: Implications for job redesign', *Administrative Science Quarterly,* Vol. 24, pp 285-308	481
	Wall T D, Jackson P R and Mullarkey S (1995), 'Further evidence on some new measures of job control, cognitive demand and production responsibility', *Journal of Organizational Behavior,* Vol. 16, 5, pp 431-435	11
	Jackson P R, Wall T D, Martin R and Davids K (1993), 'New measures of job control, cognitive demand and production responsibility', *Journal of Applied Psychology*, Vol. 78, 5, pp 753-762	24
Measures of Role Stressors	House R J and Rizzo R (1972a), 'Towards the measurement of organizational practices: Scale development and validation', *Journal of Applied Psychology,* Vol. 56, 5, pp 388-396	21

Measure	Original Reference(s)	Hits (all journals)
	House R J and Rizzo R (1972b), 'Role conflict and ambiguity as critical variables in a model of organizational Behavior', *Organizational Behavior and Human Decision Processes,* Vol. 7, 3, pp 467-505	89
Michigan Stress Assessment	French J R P and Kahn R L (1962), 'A programmatic approach to studying the industrial environment and mental health', *Journal of Social Issues,* Vol. 18 pp 1-47	23
NHS Workforce Initiative	Haynes C E, Wall R D, Bolden R I, Stride C, Rick J (1999), 'Measures of Perceived Work Characteristics for Health Services Research: Test of a Measurement Model and Normative Data', *British Journal of Health Psychology*, Vol. 4, pp 257-275	*
NIOSH Generic Job Stress Questionnaire	Hurrell J J and McLaney M A (1988), 'Exposure to Job Stress: A new psychometric instrument', *Scandinavian Journal of Work, Environment and Health,* Vol. 14, pp 27-28	20
	National Institute for Occupational Safety and Health (1997), *NIOSH Generic Job Stress Questionnaire*, Cincinnati, NIOSH	0
Objective Work Characteristics	Stansfeld S A, North F M, White I and Marmot M G (1995), 'Work characteristics and psychiatric disorder in civil servants in London', *Journal of Epidemiology and Community Health,* Vol. 49, 1, pp 48-53	20
Occupational Pressure Inventory	Not traceable	N/A
Organisational Stress Health Audit		0
Occupational Stress Indicator	Cooper C L, Sloan S J and Williams S (1988), *Occupational Stress Indicator Management Guide,* Oxford, England, NFER-Nelson	103
Occupational Stress Inventory	Osipow S H and Davis A S (1988), 'The relationship of coping resources to occupational stress and strain', *Journal of Vocational Behaviour,* Vol. 32, pp 1-15	17
Organisational Constraints Scale	Peters L H and O'Connor E J (1980), 'Situational constraints and work outcomes: The influences of a frequently overlooked construct', *Academy of Management Review, 5, 391-397*	53
Pressure Management Indicator	Williams A and Cooper C L (1998), 'Measuring occupational stress: Development of the Pressure Management Indicator', *Journal of Occupational Health Psychology*, Vol. 3(4): 306-321	2
Quality of Employment Survey	Margolis B L, Kroes W H and Quinn R R (1974), 'Job Stress: An unlisted occupational hazard', *Journal of Occupational Medicine,* Vol. 16, pp 659-661	42
	Quinn R P and Shepard L J (1974), *The 1972-1973 Quality of Employment Survey*, Ann Arbor, MI, Survey Research Centre	135
Stress Diagnostic Survey	Ivancevitch J M and Matteson M T (1980), *Stress at work,* Glenview, IL, Scott Fresman	2

Measure	Original Reference(s)	Hits (all journals)
The Job Diagnostic Survey	Hackman J R and Oldham G R (1975), 'Development of the Job Diagnostic Survey', *Journal of Applied Psychology,* Vol. 60, 2, pp 159-170	391
The Stress Profile — Derogatis	Derogatis L R (1984), 'Derogatis Stress Profile'	12
	Derogatis L R (1978), *Psychological Medicine,* Vol. 8, pp 605	1
The Stress Profile — Setterlind and Larson	Setterlind S and Larson G (1995), 'The Stress Profile – A psychosocial approach to measuring stress', *Stress Medicine,* Vol. 11, 2, pp 85-92	6
Whitehall II studies	Stansfield S A, North F M, White I and Marmot M G (1995), 'Work Characteristics and Psychiatric Disorder in Civil Servants in London', Journal of Epidemiology and Community Health, Vol. 49, pp 48-53	*
Work Autonomy Scales	Breaugh J A (1985), 'The measurement of work autonomy', *Human Relations*, Vol. 38, 6, pp 551-570	30
Work Environment Scale	Moos R H (1981), *Work environment scale manual,* Palo Alto, CA, Consulting Psychologists Press	43

* not included in original search

Source: IES, 2000

3.6 Final search results

Once the list of measures and approaches had been rationalised, and duplicate references deleted, full copies of the articles were obtained (see Table 3.4 below).

Papers which reported on measures and approaches that shared a theoretical underpinning were grouped and sent to the same reviewers.

The combined search results suggested that some measures were much more widely used in that the bulk of the papers (approximately 85 per cent) were related to just seven measures/ approaches. These measures became the 'main measures', *ie* the most widely cited/referenced measures in current use.

3.7 Review procedure

The main measures were divided among the staff available for reviewing papers. Two reviewers were allocated to each set of papers which used a particular measures or approach. In each case one reviewer was an employee of the Institute for Employment Studies, and the other was allocated from the academic collaborators. A copy of each paper was sent to the two reviewers simultaneously. Some of the reviewers had previously been involved in the development of measures. Where this was the case,

Table 3.4: Number of articles for review

Measures	No of articles
The main measures	
Job Diagnostic Survey	12
Job Stress Survey	4
Karasek Measures of Demand and Control/ Job Content Questionnaire	27
Occupational Stress Indicator	27
Rizzo and House's Measures/ Role Ambiguity and Role Conflict	44
Whitehall II Scales	15
Total number of papers for full review	**129**
All other approaches and measures	30
Total	**159**

Source: IES

it was ensured that papers were passed to other reviewers for assessment.

3.7.1 Inclusion criteria revisited

The first stage of assessing papers involved confirming, from the full paper and not just the abstract, that the work met the inclusion criteria.

A range of reasons for excluding papers for review became apparent during this process. Most striking was the extent to which researchers adapted and changed original measures. Where this was done it meant that the paper could not be used to provide evidence about the reliability and validity of the original measure. However, where changes to the measure were part of a deliberate attempt to improve or develop the measure the paper was included. This led to further rationalisation of the main measures as the Whitehall II Studies use scales based on Karasek and Effort-Reward Imbalance describes an approach not associated with a specific set of measures. This left five main measures (or groups of measures) for review.

The reasons for excluding papers at this stage (in addition to failing to meet the criteria for inclusion discussed earlier) were as follows:

1. The authors used a different response scale from that used in the original measure.
2. The authors did not use all the items which were in the original measure.

3. The authors added items to scales or sub-scales which were not in the original.
4. The authors changed the wording of some or all of the items.
5. The authors combined sub-scales for analysis which were separate in the original measure.
6. The authors scored the items differently (*eg* collapsing categories).
7. Relevant information was not provided (*eg* response rate, alpha coefficients).

Remaining papers were then assessed against recognised psychometric standards. Full technical details of the review criteria are appended, the following section provides a basic overview of the psychometric rationale for the review.

3.8 A brief overview of the psychometric criteria

Psychometrics is the branch of the psychological sciences concerned with the measurement of psychological and social issues. Psychometric analysis is therefore applicable to psychosocial hazard assessment measures. Psychometric analysis can help answer two major questions:

1) How reliable is an instrument? That is, does the instrument produce consistent measurements close up?

2) How valid is an instrument? That is, does the instrument assess what it is supposed to?

A careful look at reliability and validity indicates that they are not the same things, although they are related. It is possible for an instrument to be reliable, but not valid. For example, a watch that is consistently five minutes fast is reliable — it will indicate 12.05 pm every day at midday. But the watch is not valid — it is five minutes out. Whilst it is possible for an instrument to be consistent (reliable) but not accurate (valid), it is not possible for an instrument to be valid if it is not also reliable. In terms of psychosocial hazard assessment, the assessment should, for example, identify the same hazards for the same person doing the same job over time, provided that job does not change (that is, it is reliable) and the assessment should identify factors associated with the job that have the potential to cause some harm (that it is valid). There are numerous kinds of reliability and validity which are discussed below and in the technical appendix.

A good deal of the information needed to determine reliability and validity is statistical, and a technical appendix outlines the statistical procedures used in this review. However, in general terms, we can identify and describe several forms of reliability and validity, without having to go in to statistical theory.

3.8.1 Assessing reliability

A key part of many psychosocial hazard assessments is a self-report questionnaire, where job holders answer a number of structured questions about their job conditions on numerical scales (for example a 1-7 scale). There are two main ways of assessing reliability for such self-report instruments: (I) *internal consistency reliability* and (II) *test-retest reliability*. For instruments completed by external observers (such as consultants or members of a research team), a third form of reliability — *inter-rater reliability* — is appropriate.

(I) *Internal consistency* is essentially an index of consistency of responses to items assessing much the same thing. For example, items such as:

'I work very hard.'

and

'I have a lot of work to do.'

arguably are assessing components of work load. If the instrument is reliable, then a person answering these two items and other related items should give more or less the same answer to the questions.

(II) *Test-retest reliability* is the extent to which an instrument produces consistent measurements at two separate points in time, *provided the job does not change*. However, an instrument should also be sensitive to change: an instrument that produces the same measurements of hazards even when there has been a substantial change in the job is not sensitive and therefore not of any great use.

3.8.2 Assessing validity

There are essentially three main forms of validity: *face*, *content* and *construct* validity.

Face validity is concerned with whether an instrument looks like it measures what it should measure to a *non-expert* — for instance someone completing the assessment. Face validity can help ensure many people complete and return the instrument, as people will realise that the information may help improve working conditions. *Content* validity is where an instrument looks like it measures what it should to an *expert* — that is, someone with specialist knowledge of the nature of psychosocial hazards, such as an occupational psychologist or occupational physician. Content validity is important, because experts can help us determine whether an instrument covers the full range of relevant issues. For example, a tool that is supposed to provide a comprehensive hazard assessment would not have high content validity if key psychosocial hazard were missing from the

assessment. In addition, if there were no obvious links between the assessment and a clear theoretical framework which attempted to explain how the hazards may cause harm, questions over the content validity may also be raised.

However, it is not enough for an instrument to *look* like it measures psychosocial hazards, it is important to have some external and objective standards. We tend to seek these external and objective standards through statistical analyses to help us determine *construct* validity. An instrument has construct validity where it behaves in a way that could be predicted by underlying theory. This is done through three forms of statistical analysis:

Structural analysis: If an instrument has construct validity, then the items should assess key aspects of psychosocial hazard in coherent ways. That is, if an instrument purports to measure work demands, lack of control and lack of support and it is theorised that these three aspects of work are in some way distinct — items measuring demands should produce similar answers to each other; items measuring control should produce similar answers to each other; and items measuring support should produce similar answers to each other. If this is the case, then the instrument's 'structure' conforms to theoretical expectations.

Concurrent analysis: If an instrument is valid, it should be associated with measures of harm assessed at the same time (concurrently) as the psychosocial hazard assessment (assuming that the hazards have been present in the working environment for a sufficient time to bring about harm). For instance, jobs identified as having high levels of psychosocial hazards should also be associated with high levels of depression, anxiety *etc.*, if the assessment method is valid. There should also be consistent differences in harm levels between those engaged in jobs that could reasonably be expected to differ on psychosocial hazards.

Predictive analysis: Predictive analysis is like concurrent analysis — but harm is measured some time after the assessment of psychosocial hazards. That is, a valid instrument should be able to predict future harm. This indicates that the instrument is capable of detecting exposure to hazards that can cause harm. While all the kinds of validity and reliability discussed here are important it should be noted that the predictive validity of measures is perhaps the single most important kind of validity. If measures of hazards do not predict future harms they cannot be considered to be valid measures of hazards (unless there are very good reasons why this is the case). In other words, the entire purpose of measuring hazards is because they are assumed to be related to, and can cause future harm. If measures of hazards are unrelated to future harm they have no place within a risk assessment framework.

Discriminant analysis: If an instrument is valid, then it should *not* be related to irrelevant issues. That is, the instrument should

measure something to do with work-related hazards, not, for example, aspects of personality. This is discriminant validity.

3.9 Development of a pro forma to capture assessments

Some of the measures used were single scales whilst others actually comprised several sub-scales. In addition, different studies tend to use these scales in different ways. For example, one study may use all the sub-scales in a measure while another may use only one. To ensure all the relevant evidence about reliability and validity was recorded, a form was used by each reviewer to report and structure the relevant information for each scale or sub-scale used in the study.

Separate pro formas were developed to capture the detail of any exploratory and/or confirmatory factor analyses.

Example pro formas are given in Appendix 2.

3.10 Assessment criteria and information sought

In summary, the review sought to address the following points about each instrument:

Sensible questions in each measure, derived from strong theory; qualitative analyses; or both. The questions should be clear, unambiguous and reflect the constructs they are assessing according to experts — this is content validity. Face validity is desirable.

Sensible response format for the items: preferably a clear and recent time frame; not a Likert scale format (*ie* strongly agree to disagree — as such scales might be more reflective of attitudes to the work rather than necessarily the experience of events at work); frequency based rating if the format is meant to measure incidence (*eg* not at all, once or twice, sometimes); another sensible rating format.

A clear factor structure: preferably replicated across samples (and studies), and preferably at some point subject to confirmatory factor analysis.

A reliable instrument: first internal consistency; test-retest where work conditions are stable; low test-retest where work conditions are expected to change over time.

Concurrent and predictive validity.

High response rates from surveys using the instrument.

Discriminant validity, and the extent to which these measures are relatively 'free' from contamination from personality variables such as disposition towards negative emotions (negative affectivity).

Information about design flaws or other aspects of the study which are relevant to assessing the reliability and validity of the measures.

4. Review of Main Measures

This chapter focuses on the main measures of psychosocial hazards for which reasonable evidence of their reliability and validity was available in the literature. It starts with general findings about the sources of data used for the review, where the information comes from and how the field of research into organisational stress has developed as a whole. It provides an overview of the types of issues that are currently a source of debate and provide an important context for how and why measures have developed in the way that we find them today. It then goes on to discuss some of the key issues about validity which apply to all the measures reviewed.

The final sections are devoted to a review of the evidence available on different measures of psychosocial hazards. Each section gives a broad introduction to the measure, clarifying what the measure is supposed to do and how widely it is used. Evidence about its validity and reliability is presented. Where available, additional information is provided on its development, and broader aims and purposes. The extent of available evidence used to review each measure is also made clear.

4.1 Introduction to, and overview of, the sources of evidence used

As indicated elsewhere, most of the evidence for the reliability and validity of the hazard measures reviewed in this report was extracted from research papers published in peer-reviewed academic journals in a number of different fields. It is envisaged that relatively few readers of this report will be familiar with the research publications field, the areas in which these papers are published, or the nature of the available measures of psychosocial hazards. There are a number of important features of these journals and stress research field in general that it is helpful to clarify so that the nature of this evidence can be more fully understood. Each of these areas will be discussed in turn.

4.1.1 Peer-reviewed journals

We chose to focus on this type of journal as the quality of research they publish is likely to be higher than that found in other sources. This is for two related reasons. First, these journals are generally regarded as the most prestigious and most researchers would aim to get their work published in such journals rather then other sources. The second, related point, is that only papers which have undergone a strict review process, and been deemed acceptable by both the reviewers and the editors, are accepted.

For higher status peer-reviewed journals, the process would broadly be as follows:

- Submission of manuscript.
- The editor decides whether to send it for review.
- Expert reviewers (usually at least three) read and comment on the article (which is blind reviewed and the identities of the reviewers are likewise not known to the authors).
- Comments and recommendations are given to the editor.
- Editor makes a decision (reject, accept, revise and resubmit).

If revise and resubmit, then the process continues:

- Submission of revised manuscript.
- Back to original reviewers.
- More comments.
- Editor makes decision on revised manuscript (reject, accept, revise and resubmit).

In some cases, further iterations will take place.

Of course, this review process is by no means a guarantee of quality, and the standard of reviewing varies enormously across peer reviewed journals. However, it is still reasonable to suggest that research published in these journals is likely to be of a higher quality. The focus on peer-reviewed journals does not imply that high quality research is only to be found in such journals — but rather, that within the public domain, it is the most likely source.

4.2 Overview of the occupational stress research field

While these journals come from a number of different discipline areas such as occupational medicine, organisational psychology, and management, most of the papers we have reviewed can be viewed as falling within the multi-disciplinary research field usually described as organisational or occupational stress. While this report uses the terminology of psychosocial hazards these are in effect what are also called stressors in the organisational stress field.

4.2.1 An interdisciplinary field

How can this field best be characterised? As already mentioned, it is interdisciplinary. However, it does appear that psychology and the behavioural sciences tend to dominate. This means that there is much focus on the ways in which employees *perceive* their workplace and the potential hazards within it, and also the ways in which these *perceptions* may in turn impact on employees' psychological well-being. This perspective has important implications for the kinds of hazard measures that have been developed, as will be discussed below.

4.2.2 Definitional and conceptual issues

Another way of describing the field is that it always has been and remains engaged in quite fundamental debates about both the definitions of stress (*ie* what does stress actually mean?) and the theoretical bases of stress (*ie* how can we understand and explain what stress is and what it is supposed to do?). While it is common for many fields to engage in such debates, the fundamental and on-going nature of these make the organisational stress field somewhat unusual. This feature of the field also has implications for how hazards are measured.

4.2.3 Methodological issues

Finally, the organisational stress field has always and continues to grapple with methodological issues (*ie* how can stress be measured? What are the best ways to design studies to demonstrate the causes and consequences of stress?). This is in some ways unsurprising given the definitional and conceptual issues described above: if we are unsure what we mean by stress or how it might work, researching it is likely to present problems. The two methodological issues most relevant to this report are the assessment of stressors (hazards) and the identification of cause and effect.

There has been considerable debate about whether hazards should be measured objectively or subjectively, or perhaps in both ways. On the one hand, it is clear that employees' perceptions of hazards are central: an employee must perceive a psychosocial hazard otherwise it is not likely that it can cause harm. On the other hand, there must also be some objective basis to a hazard: it cannot all simply be a matter of employee perception. This debate is very much alive and was covered recently in a special issue of *Journal of Organizational Behavior* (Vol. 20, 1999). A further complicating factor in the assessment of stressors is the nature of the role of individual differences. While it is reasonably clear that aspects of people's make-up, such as personality, do influence the extent to which they both perceive and react to stressors, it remains unclear how such individual differences work or how, and indeed if, they should be incorporated into research to control

for their effects. This issue has likewise been debated at some length in the research literature.

The second major methodological issue is the identification of cause and effect. While detecting and understanding cause and effect can be complex, at the very simplest level it requires longitudinal studies where the cause (in this case hazards) is measured some time before the effects (in this case harms). The vast majority of studies in the organisational stress field are cross-sectional (*ie* a one-off study where everything is measured at the same time) which unfortunately can tell us nothing about cause and effect. This is obviously a very serious shortcoming in a field whose main objective is to understand how work stressors may be a cause of reduced well-being. This limitation is regarded as so serious that over the past decade a number of important journals (for example, *Journal of Applied Psychology*, *Journal of Occupational and Organizational Psychology*, *Human Relations*) actively discourage the submission of papers which report stress studies using such cross-sectional designs. This aspect of the field likewise has implications for the available evidence.

Last, most of the published studies in the field fall into one of a number of categories, including the following. The simplest form of study will simply describe the hazards and/or harms experienced by a particular occupational group. A second type of study aims to examine associations between measures of hazards and measures of harms to assess which kinds of hazards appear to be most strongly associated with harm. A third kind attempts to look at the numerous factors involved in hazard-harm relationships, such as personality, coping skills, and perhaps other aspects of the work environment. For example, a study may attempt to see whether aspects of employees' personalities increase or reduce the strength of the relationships between hazards and harms. Finally, there are also many methodological studies that aim to address some of the methodological issues outlined above.

It should be noted that there are relatively few studies which have as their primary goal the psychometric evaluation of hazard measures, and hence most of the evidence cited here comes indirectly from papers which report studies that have used relevant measures and which contain data that tell us something about reliability and validity.

While there is very broad agreement in this field on what 'stress' in the most very general sense may mean, definitional, conceptual, and methodological issues and debate have tended to dominate. There is little doubt that psychosocial hazards do cause harm to employees, but the field remains some way off a sound understanding of how, why and the extent to which this happens.

4.3 Typical hazard measures

Most measures of psychosocial hazards have similar features. First, they are self-report and perceptual (*ie* not based on observation or the measurement of some aspect of work). They tend to contain lists of statements (items) describing aspects of work that may represent psychosocial hazards (for example: 'I have little control over the way my work is scheduled', 'I often experience marked increases in workload'). Respondents are then required to respond to the statement usually by indicating on a scale (which is then given a numerical value) the extent to which they agree or disagree with the statement or how often they experience the situation described. Usually, for reasons of reliability and validity, a particular hazard such as workload will be assessed by a number of similar items (a scale or sub-scale) and the total of, or average score on, those items used as an indicator of the level or amount of that hazard perceived by the respondent.

A second feature of typical hazard measures is that they often do not appear to draw heavily on well-established theories or indeed on any theory at all. While there are numerous frameworks that provide quite comprehensive overviews of possible hazards, harms, and the relationships between the two, there are relatively few theories about the nature of stress which can be used to guide the measurement of hazards. It may appear to those previously unfamiliar with hazard measures that many of those reviewed in this report appear to owe more to 'common sense' than to sound theory — in many cases this may be true. Such measures can sometimes seem to be little more that lists of things that may cause people difficulties at work that have been grouped together in various ways.

A third feature of hazard measures is that many were developed as research tools for assessing specific job characteristics (for example: demands, control, conflict) that, as it happens, can also be considered to be psychosocial hazards. Many were not developed specifically as measures of hazards nor for use in organisations as a practical hazard assessment tool. What these measures offer and what is required for practical hazard assessment may not therefore be one and the same thing. This has considerable implications for this study when considering certain aspects of validity and in particular when contemplating the theory or explanation of how these psychosocial hazards cause harm. The next section considers some of the broad issues about theory and its role in the validity of the measures being reviewed.

4.4 Theory and validity

Almost all of the evidence discussed here about the reliability and validity of the psychosocial hazard measures under review comes from considering their statistical properties — how they 'perform'

empirically when used in studies (see Section 3.9 for an overview or Appendix I for the detail). There are, however, some aspects of validity that are unrelated to the psychometric properties of the measure itself and instead are related more fundamentally to the theory on which the measure is based. Although a detailed discussion of the theoretical aspects of validity is outside the scope of this review, they are, nonetheless important and form the basis for considering nearly all aspects of validity.

In essence, theory is the foundation on which all measurement is based. Without sound theory, the meaning of any findings based on a measure remain obscure and, most importantly in this context, the practical relevance of the finding is difficult to establish.

Theory is vital as it attempts to answer the questions of why and how things work or may work. For example, precisely how a hazard causes harm is a theoretical question even though we may have data which clearly show such a relationship exists. To take a more specific example, the reasons why low control may lead to harm is a theoretical question. The answers to such theoretical questions are also important for practical reasons, as without knowing how something works or might operate, it is very difficult to intervene in any systematic or strategic way. All we know is that there is a problem not how or why it has come about, nor what we can do to solve it.

As indicated earlier, the organisational stress field has numerous models which describe, but by definition and purpose, do not explain the possible relationships between hazards and harms. These models are not theories as they cannot and do not address questions about why and how hazard-harm relationships exist. In general, there are relatively few well-developed theories within the organisational stress field that can be used as a foundation for the development of psychosocial hazard measures.

One of the few exceptions to this is the relatively recent effort-reward imbalance model developed by Siegrist and a colleague (Siegrist and Peter, 1996) in which theory does appear to drive measurement in a direct way. Other approaches which may be thought of as theoretical, such as Karasek's (1979) Job Demand-Job Control Model, though they do focus measurement on particular kinds of hazards, do not in themselves suggest ways in which measures can or should be developed.

Given that most of the measures we will review share a similar theoretical basis in that they are not derived from strong theory, it is possible to discuss in general the extent to which the validity of these measures is compromised by the limited development and use of theory in psychosocial hazard measurement. Other aspects of validity will be discussed for each measure separately later. Two kinds of validity, content and construct, will be discussed in turn.

4.4.1 Theory and content validity

As discussed earlier, a key aspect of content validity is whether or not the measure looks like it measures what theory suggest it should measure. Another aspect of content validity is whether or not the response formats make sense, given the theoretical bases of the phenomenon under investigation. Given the low level of theory adopted in psychosocial hazard measure development it is not always clear precisely what the measure should or should not cover. The explanation of what the hazard is and how or why it may have its effects is often unspecified, which means it is not possible to determine what the measure should cover. A hazard such as workload, for example, is often conceptualised somewhat weakly around broad notions of 'demands' or having 'too much to do' which is then reflected in workload measures which can seem somewhat non-specific and unfocused. It is possible, for example, to theorise about many different types of workload and kinds of demands and many different types of effect, but this has not happened to any significant extent.

Likewise, response formats should ideally be based on theory about the nature of the particular hazard under investigation. For example, an 'agree-disagree' response format implies that we are measuring something akin to an attitude, whereas a frequency format such as 'none of the time — every day' format implies an event-based kinds of hazards which respondents are able to recall and report. Again, the very limited use of theory inevitably means that response formats are not based on theoretical assumptions.

The content validity of most of the measures reviewed here, as evaluated by an examination of underlying theory, is therefore somewhat low simply because the theory used, if any, does not make clear specifications about what the measure should include or how the response format should be designed.

4.4.2 Theory and construct validity

Broadly speaking, construct validity is concerned with whether a measure behaves empirically in ways which would be predicted by underlying theory in terms of its structure and relationships with other measures. In order therefore to provide a thorough assessment of construct validity, it is vital that underlying theory specifies the kinds of relationships we should expect to find and why we should expect to find them.

Unfortunately, given the limitations of underlying theory, it is not usually possible to specify precisely what kinds of relationships we might expect which has major implications for assessing at least three kinds of construct validity. Concurrent validity requires that we specify what kinds of associations should exist between the measure and other theoretically related measures. Discriminant validity, on the other hand, is determined by

checking whether there are not significant relationships between the measure and variables which are theoretically unrelated to the measure. Third, predictive validity means examining expected relationships between hazards and future harms. In each case, this requires a good level of specification of what the hazard is, how it works, what is should and should not be related to, and, most importantly in this context, which particular harms may be caused or predicted by the presence of the hazard.

The construct validity of many of the measures in this review is low simply because the theory, where any is used, is at such a general or descriptive level that we cannot identify the specific relationships we would expect to find between the measure and other measures.

The next sections look in more detail at the five main measures identified for inclusion in the review.

4.5 Job Diagnostic Survey (JDS)

The Job Diagnostic Survey was developed by Hackman and Oldham in 1975 as part of a study of jobs and how people react to them. It was designed to explore the Job Characteristics Model, in which perceived job characteristics can cause affective responses to the work environment. Its aim is also to help determine how jobs can be better designed on the basis of how individuals react to different types of job. It has been described as one of the principal self report measures for assessing work characteristics (Fried, 1991).

The JDS comprises seven main scales which assess the following work characteristics: skill variety; autonomy; task identity; task significance; job feedback; feedback from others; and, dealing with others. Of these, most attention has focussed on the first five scales; the latter two scales (covering relations with others) often being dropped from research studies. Two forms of the original questionnaire exist: the 'short form' consisting of 53 items and a longer version. A revised version has also been proposed.

This research identified 14 relevant papers for review through the literature search. However, eight papers had to be dropped on subsequent reading of the full paper, and one was a review, leaving five original research studies from which evidence is drawn.

Despite being a well known measure, and having been widely used in the 1970s and 1980s, the JDS has received little attention over the last decade.

4.5.1 Reliability

Only three papers reported original reliability evidence for the main five scales (Munz, Huelsman, Konold, McKinney, 1996; Spector and Jex, 1991; and Champoux, 1991). This is presented in Table 4.1 and shows that internal consistency of the sub-scales fluctuates from study to study, but is moderate on most scales, with the exception of Task Significance. A major review of the JDS (Taber and Taylor, 1990) has pointed to some evidence that the JDS has an unstable factor structure and low internal consistencies on some scales.

Reliabilities for scales in the revised JDS were found to be a general improvement on the original version of the questionnaire (Spector and Jex, 1991; Cordery and Sevastos, 1993).

No evidence was identified for test-retest reliability or test-retest sensitivity.

One paper reported on inter-rater reliability and found good evidence for consistency across raters.

4.5.2 Validity

Face validity

Early work by Hackman and Oldman (1975) points to good face validity for this measure with items being checked and re-checked over a two year developmental period.

Content validity

Content validity appears reasonable for the JDS in relation to item content. Hackman and Oldham propose that the motivating potential of a job called the 'MPS Score' is based on the experienced meaningfulness of the work (skill variety, task identity, task significance), the degree of autonomy and the degree of feedback. Early work on the scale involved three major revisions to hone and refine item content based both on psychometric qualities and substantive considerations. However,

Table 4.1: Scale reliabilities for the Job Diagnostic Survey

Name of scale	Highest Alpha	Lowest Alpha	Mean Alpha
Skill variety	0.78	0.70	0.73
Autonomy	0.87	0.69	0.75
Task identity	0.81	0.64	0.71
Task significance	0.74	0.54	0.64
Job feedback	0.83	0.64	0.72

Source: IES, 2000

the scales are not timebound and the wording is somewhat old fashioned in places.

Concurrent validity

Concurrent validity data shows reasonable correlations with job satisfaction, and good correlations with other measures of psychosocial hazards.

Predictive validity

Predictive validity was not found to be good on the original measure and the revised questionnaire demonstrated little improvement.

Discriminant validity

Discriminant validity for the scales in the JDS is moderate. Correlation matrices from two studies show that the sub-scales are inter-correlated and to some degree correlate with theoretically unrelated variables (*eg* social satisfaction). However, one study examining positive and negative affect (Munz, Huelsman, Konold and McKinney, 1996) suggest negative affect is not associated with the relationships between these scales.

4.5.3 Utility

The JDS is interesting in construction as it attempts to obtain a 'norm' or objective comparison from respondents for each of the sub-scales. Each respondent is asked first to make an assessment of the extent to which their job involves certain characteristics (*eg* allows them autonomy to decide how and when they do things). They are then asked about how the different aspects of their job makes them feel.

Example items include:

> *'To what extent does your job require you to work with mechanical equipment?'*

To which respondents are asked to give an objective/accurate answer.

This is then followed by a series of questions or statements where respondents are asked to:

- rate the accuracy of a statement (*eg* the job requires a lot of co-operative work with other people); and
- indicate how much they agree with a statement (*eg* my opinion of myself goes up when I do this job well).

The measure is freely available in the literature and has a scoring key for both the short and long versions. It is generally applicable across industrial sectors and occupations. Some limited norm data are available in the literature.

4.6 Job Stress Survey (JSS)

The Job Stress Survey was developed by Spielberger from his earlier work with law enforcement officers and teachers. It is a relatively new measure, the *Professional Manual for the Job Stress Survey* being published in 1994. Both its predecessors, the 'Police Stress Survey' and the 'Teachers Stress Survey', were designed in a deliberate attempt to try and address some of the criticisms levelled at existing measures of stress. Specifically and importantly they asked not just about the severity of an incident, but also about the frequency with which the incident/experience occurred. The JSS follows the same format as these two earlier surveys and is described by Spielberger and Reheiser (1995) in the following way:

> *'This 30 item psychometric instrument was designed to assess the perceived intensity (severity) and frequency of occurrence of working conditions that are likely to adversely affect the psychological well-being of employees who are exposed to them.'*

The JSS provides overall scores on severity of stressful experience (all 30 items) and frequency of stressful experience (30 items) and a Job Stress Index (the sum of the cross products of the severity and frequency scores). There are also job pressure and organisational support sub-scales (both ten items) for which severity and frequency scores can be computed.

The items were selected to describe generic sources of stress in a range of occupational settings for managerial, professional and clerical employees.

This research identified six relevant papers for review through the literature search. However, three papers had to be dropped on subsequent reading of the full paper, leaving three original research studies from which evidence is drawn.

4.6.1 Reliability

A thorough review of the literature has revealed that limited reliability data are available on this measure. Two of the three papers provide reliability data on the scales, one reports on data from several studies and is not considered here to avoid double counting. A large scale study (N = 2,389) reports reliabilities for the severity, frequency and index measures in three contrasting occupational groups: university; corporate; and military employees. Reliabilities for all scales are high, suggesting a good level of internal consistency. Details are given in Table 4.2.

No data are given for the sub-scales, nor is data given on test-retest reliability, test-retest sensitivity or inter-rater reliability.

4.6.2 Validity

Face validity

The development of the JSS suggests that it probably has reasonable face validity. Items were drawn from existing measures (namely the 'Police Stress Survey' and the 'Teacher Stress Survey') and adapted to be appropriate for use in a wide range of settings. However, from the papers included in this review, there is no evidence of any attempt to develop or pre-test JSS items on a target population.

Content validity

The content validity of the JSS is best described as good to marginal using the psychometric criteria detailed in Appendix 1. The measure uses a frequency based response scale with a specified time period. However, as previously noted, the measure uses items drawn from scales specific to police officers and teachers and there is no evidence of testing the items on the (more general) target population. It is therefore difficult to assess the extent to which it covers the full range of relevant phenomena.

Construct validity

Construct validity concerns the structure of a measure and whether or not it behaves in a way that could be predicted from the theory. One of the three studies includes an exploratory factor analysis of the JSS — this looks at whether or not people respond as we would expect and if those relationships hold true for different groups. For the JSS the results were marginal to good, with the data supporting the proposed structure of two distinct sub-scales measuring job pressure and organisational support.

Table 4:2: Scale reliabilities for the Job Stress Survey

Sector	Gender	Severity Scale	Frequency Scale	Index
University	Male	.91	.89	.89
	Female	.93	.92	.93
Corporate	Male	.88	.88	.85
	Female	.90	.89	.88
Military	Male	.84	.82	.79
	Female	.81	.74	.71

Source: Spielberger and Reheiser, 1991

Concurrent validity

Only one study looked at whether or not the sub-scales had the expected relationships with other measures and behaved in the expected way – in other words, are the JSS sub-scales related to other measures as we would expect (*eg* are high levels of stress related to high levels of psychological harm?). The findings from the available research on the JSS were somewhat mixed.

Predictive validity

No data were found on predictive validity.

Discriminant validity

Discriminant validity data for this measure were not reported and could not be ascertained from the papers reviewed.

4.6.3 Utility

The JSS marks an important development in the design of hazard measures. It is constructed to ask two separate questions of the respondent: How **much** is it a problem? and: How **often** is it a problem? Understanding the frequency of an event or experience is an important aspect of any assessment of hazards, and is an aspect of measurement overlooked by the majority of other measures.

That said, the JSS also has a number of weaknesses. To date, there have been relatively little published data on reliability and validity which makes it hard to fully assess the strength of the measure. What limited work has been published is good, but ideally these findings should be replicated across a range of different settings and occupational groups before firm conclusions about the performance of the measure can be drawn. Additionally, while the items in the scale (face and content validity) seem reasonable, and the structure of the scale (construct validity) appears good, evidence on the relationship between this measure and other theoretically related scales (concurrent validity) was mixed. There is no evidence of the ability of this measure to predict whether or not subsequent harm will occur (let alone the extent of such harm) following exposure to the hazards identified. Finally, the most consistent relationships with the JSS sub-scales and the JSI (Job Stress Index: see Section 4.6) were found with locus of control (LOC). As LOC is often characterised as an individual trait (a way of viewing the world around you as within or without your control) this calls into question the ability of the scale to discriminate general views from specific work-related risks.

The JSS is a 30 item measure, designed for use with managerial, professional and clerical employees. Respondents are asked to

indicate the amount of stress (*ie* the severity) they perceive to be associated with each of the items on a scale of 1–9 in comparison to a standard stressor (the assignment of disagreeable duties) which is rated at 5. Respondents are then asked to indicate (on a scale of 1–9+) the number of days over the previous six months when they have been exposed to/encountered this stressor. Typical items include 'excessive paperwork' and 'working overtime' (Spielberger and Reheiser, 1994).

The measure and a guidance manual are available commercially.

4.7 Karasek Demands and Control/Job Content Questionnaire (JCQ)

Karasek's Job Demand-Control (JD-C) Model has been arguably the most influential in stress research since it was first introduced in 1979 (Karasek, 1979). The model itself focuses on two aspects of work — the demands of the job, and the decision latitude available to the individual. Karasek, Brisson, Kawakami, Houtnam, Bongers and Amick (1998, p.322) describe the model in the following way:

> *'The most commonly used demand/control model hypothesis predicts that the most adverse reactions of psychological strain occur when the psychological demands are high and the worker's decision latitude is low.'*

Various other predictions exist about the relative importance of demands and control in predicting employee ill-health. Primary amongst these is the proposal that control moderates the effects of demand, so jobs that are high in both demand and control are not necessarily associated with poor psychological outcomes. The JD-C model has promoted large numbers of research studies and generated many measures of demands and control. As a result, the JD-C model has focussed researchers' attention on the importance of demands and control in understanding psycho-social hazards in the workplace, in particular the role of control. It can be seen in many ways more as a broad approach to the measurement of psychosocial hazards. In the 1980s a social support element was added to the model and this extension of the original is known as the Job Demands-Control-Support (JDCS) model. JD-C has also been referred to as the Job Strain Model.

One specific measure, the Job Content Questionnaire (JCQ), was developed by Karasek and published in 1985. This review is primarily concerned with the JCQ and the two scales most commonly used from it (demands and control).

The JCQ consists of five main scales:

- Decision Latitude/Job control (comprising skill discretion and decision authority).

- Psychological Demands.
- Social Support.
- Physical Demands.
- Job Insecurity.

As has been indicated above, the JD-C model has been a very popular basis on which to approach stressor research. As a result we were able to identify 34 papers which, on the basis of the abstract, looked appropriate for inclusion in the review. On closer perusal, however, only 12 fulfilled the inclusion criteria. Many of the excluded papers were directly concerned with exploring the relevance and validity of the JD-C model to current working environments. Unfortunately, for the purposes of this study, they used one-off variations or adaptations of some of the JCQ scales, so could not be used for establishing reliability or validity.

Of the remaining 12 papers, only two used all five JCQ scales. Five of the papers used just the demands and control scales; two papers used three scales (demands, control and social support); and in some of the papers reliabilities are given separately for the skill discretion and decision authority sub-scales of the decision latitude measure. Additionally, five papers were found which related to a revised version of the demands and control scales used in an advanced manufacturing technology setting (Jackson, Wall, Martin and Davids, 1993) and several papers were found which used revised scales amongst London based, civil servants (the Whitehall II Studies, see for example Standfeld, White, North and Marmot 1995). These studies are considered separately at Section 4.8.

4.7.1 Reliability

The two studies which report on all five JCQ scales do not report scale reliabilities. Three of the five studies using just the demand and control measures report whole scales reliabilities. One paper provides separate reliabilities for the decision latitude and skill discretion sub-scales, job demands and social support. Overall, the following can be summarised:

- Reliabilities for the job demands scale range from .64 to .81.
- Reliabilities for the job control scale taken as a whole are generally much better, ranging from .77 to .86.
- The one paper reporting separate reliabilities for the job control sub-scales found skill discretion scale to have a reliability of .74, where as the decision latitude scale had a reliability of .65.
- One reliability was reported for the social support scale at .81.

The original job control scale has come in for criticism in the literature for mixing together skill discretion and decision latitude. This may in part explain why so many different and adapted versions of the scale exist. It also suggests that there are probably more reliable versions of this scale available, but an investigation of all of them is beyond the remit of this particular study.

4.7.2 Validity

Face validity

Face validity appears good for the JCQ scales, with items being developed and revised over decades and adapted for the UK.

Content validity

Content validity is arguably more mixed. The JCQ assesses key areas in relation to the JD-C model. However, some researchers have found the need to adapt and develop the scales for specific settings. Additionally, there has been some suggestion that the scales would benefit from revisions to include psychosocial hazards associated with recent economic and technological changes or that the scales might need to be revised for different study populations.

Frequency based rating scales are not used, but questions are phrased in a way to avoid cognitive bias.

Taking a broad perspective, the scale measures only a small range of possible psychosocial hazards.

Construct validity

One lasting criticism of the JCQ has been the structure of the job control measure and the extent to which it accurately measures distinct components of control. This criticism has led to many different variants of the control measure being developed. On the whole though, control has been found to be a useful construct. The demand scale is less contentious.

Concurrent validity

Generally, the evidence suggests good concurrent validity for the JCQ. Reported associations indicate patterns of relationships with other variables as would be expected with significant negative relationships between decision latitude, skill discretion, decision authority and education in general. Significant positive relationships are related between psychological demands and education, but a significant negative relationship between physical demands and education.

Predictive validity

The majority of papers included in the review do not report on predictive validity for this measure. However, the four papers that do examine this area indicate that predictive validity is reasonable. All four studies report findings that indicate the measure of control predicts future health (mainly cardiovascular health).

Discriminant validity

Few papers reported or provided evidence on discriminant validity. In those that did, the findings were mixed. In some studies the expected relationships were found. In others, authors suggested that other factors such as coping styles, or socio-economic status might contribute to the observed relationship between job characteristics and strain. It suggests that for some groups of workers, in certain settings, factors other than those assessed by this measure are important in determining reactions to the working environment.

4.7.3 Utility

The JCQ consists of 49 questions across five scales. It is widely applicable across different sectors and jobs and has been used extensively with different occupational groups.

Extensive data are available via the JCQ Centre and the JCQ users network. The following is the JCQ Usage Policy:

> *'The JCQ is copyrighted and not published in the public domain; however, it is the goal of the JCQ centre to make it available to all researchers who request it with substantial supporting documentation, and to promote scientific development in the area through a users' network. The JCQ Questionnaire and users' guide and research documentation are provided free of charge to most users. However, JCQ use by large research studies (over 750 participants) and commercial users requires payment of per use charges. Registration in a JCQ users' project database for the users' network for all users and a copy of the researchers JCQ and demographic data for future reliability analysis (large studies only) are required. Contact the JCQ Center, Department of Work Environment, University of Massachusetts Lowell, Lowell, Massachusetts 01854, for details of policy fees and requirements.'*

4.8 Other measures of demand and control

Although there are many one off versions of demands and control measures based on Karasek's job strain model and the JCQ, two groups of studies emerge from the literature which offer significant developments of the demands and control scales.

The first group of studies centres on work by Jackson and colleagues to develop more specific measures of demands and control for use in an advanced manufacturing setting.

The second group of papers stem from a major longitudinal research programme which has been underway since 1985, involving several thousand London based civil servants. Known as the Whitehall II study, the work has involved many researchers, but has been led predominantly by Marmot and Stansfeld. This research has also used measures of work characteristics based on Karasek's job strain model and the JCQ.

Both groups of studies are discussed in turn.

4.8.1 Jackson, Wall, Martin and Davids: measures of demands and control

In 1993, Jackson, Wall, Martin and Davids first reported on work to develop:

> *'…standardised and widely applicable measures…(to) allow the accumulation of comparative and normative data that is necessary to make more systematic judgements about whether job demands are at critical levels.'*

Their research was based on samples from an advanced manufacturing technology setting and sought to develop far more specific measures of demands and control.

Items for the research were developed from interviews and existing measures. They were scrutinised for complexity, ambiguousness or duplication, and 22 items were finally selected. A five point response scale was used which ranged from 'not at all' to 'a great deal'. Statistical analysis of the data gathered from two samples supported a five factor/sub-scale structure — timing control, including such items as: 'Do you set your own pace of work?' and: 'Do you decide the order in which you do things?'; method control, with items such as: 'Can you vary how you do your work?' and: 'Can you control the quality of what you produce?'; monitoring demand, for example: 'Do you have to keep track of more than one process at a time?'; problem solving demand, *eg*: 'Do you come across problems in your job that you have not met before?'; and production responsibility which included items such as: 'Could your alertness prevent a costly loss of output?'. Reliabilities for the scales are given below.

A further study using just the timing and method control scales found reliabilities of .75 for timing control and .69 for method control. One study which uses a combined method and timing control measure reported a scale reliability of .83.

Table 4.4: Jackson, Wall, Martin and Davids — Demands and Control Scales

Name of scale	Alpha S1	Alpha S2	Alpha S3
Timing control	.85	.79	.86
Method control	.77	.80	.76
Monitoring demand	.73	.75	.67
Problem solving demand	.50	.67	.60
Production responsibility	.90	.86	.85

S1 = study 1; S2 = study 2; S3 = study 3.

Source: IES, 2000

4.8.2 Validity

Face validity

Face validity is considered to be good for this measure, with items derived from interview or established measures, and scrutinised for complexity *etc.*

Content validity

Content validity appears reasonable given the manufacturing focus of the current measure. Its validity in other setting is doubtful, and as with the JCQ, given a broader setting it could be criticised for only reflecting a small range of potential psychosocial hazards.

Construct validity

Evidence presented by Jackson, Wall, Martin and Davids (1993) points to reasonable construct validity, with the factor structure being maintained over two samples. However, reliabilities for problem solving demand remain low suggesting a problem with this particular scale.

Concurrent validity

Jackson's results also suggested reasonable concurrent validity in a manufacturing setting with the scale able to distinguish accurately between supervisors and subordinates in jobs with some shared and some distinct aspects.

Predictive validity

No data on predictive validity were identified.

Discriminant validity

Further work by Mullarkey, Jackson, Wall, Wilson and Grey-Taylor (1997) on the timing and method control scales only demonstrates the relationship of the scales with different harm outcomes (and indicates different patterns of relationships for the scales).

4.8.3 Utility

The Jackson demands and control measures are in the public domain and are freely available.

4.9 The Whitehall II studies

The final version of demands and control measures covered by this review are those used in the Whitehall II studies. The Whitehall II studies refers to a major longitudinal research programme which has produced a unique data set in job stress research. Commencing in 1985, the research programme has tracked large numbers of London based civil servants, collecting both questionnaire and physiological data at several different time points. Over ten thousand civil servants across 20 departments participated in the research and as such the data set provides a rare opportunity to examine the relationship between work characteristics and physical and psychological health.

The work characteristics measure in the Whitehall II studies is described as a sixty seven item self report questionnaire covering job strain (*ie* demands and control) social support, job satisfaction and coping skills. All questions are answered on a frequency based response format – a four point scale ranging from 'often' to 'never/almost never'.

Only one set of reliability data was identified for the scales amongst the papers reviewed, details are given in Table 4.5.

Reliabilities are good with the exception of the job demands scale which falls below the acceptable threshold (of 0.7) although this might be in part due to its brevity.

Table 4.5: Scale reliabilities for the Whitehall II Studies

Scale	Reliability
Decision latitude (control – 15 items)	0.84
Job demands (four items)	0.67
Social support (six items)	0.79

Source: IES 2000

Face validity

No detail on the source or development of items was given in the papers reviewed other than they were based on the job strain model.

Content validity

Limited examples of items were cited in the papers reviewed, and as for face validity, there was no information on the development of items for the scales. That taken into account, content validity appears reasonable in so far as it follows Karasek's job strain model, includes aspect of work based social support and uses a frequency based response format.

Construct validity

Only one of the papers included in the review reports that a principal components analysis of the work characteristics scale identified seven different dimensions: 'work pace' and 'conflicting demands' (similar to Karasek's 'psychological demands'); 'skill use' and 'variety and control' (similar to Karasek's 'decision latitude') (NB additional dimensions include 'social support', 'job importance' and 'job satisfaction'). No other evidence was identified on the construct validity of the scales and it is therefore difficult to draw any firm conclusions.

Concurrent validity

On the whole, where cross sectional analyses in the reviewed studies were reported, they revealed the anticipated associations between work characteristics and theoretically related variables. For example decision authority, skill discretion and work social support were associated with low levels of anxiety and depression, whereas high job demands were associated with poor psychological health.

Concurrent validity for these measures would appear to be good.

Predictive validity

The longitudinal nature of the Whitehall II research means that, unlike the other measures reviewed here, there are a number of articles that provide data on the predictive validity of the work characteristic scales. Five papers reviewed here provide information on the predictors of subsequent physical and psychological ill health.

Overall data show that the measures of work characteristics used in the Whitehall II studies are good predictors of subsequent physical, psychological and social functioning some years on.

Importantly, the research has shown that these relationships hold true even when personality characteristics such as negative affectivity (a persons propensity to perceive and answer questions negatively regardless of the objective circumstances) are taken into account.

Discriminant validity

From the available evidence, the Whitehall II work characteristic scales also appear to have good discriminant validity. Scales appear to be associated with specific effects rather than increasing general susceptibility.

For example, the job demands scale (high scores *eg* high pace of work or conflicting demands) was found to be predictive of future psychiatric disorder whereas low job control was found to increase the risk of heart disease. Social support at work was generally found to be a positive factor in determining psychological health as was decision authority.

Utility

The Whitehall II studies use relatively short measures of decision latitude (15 items) job control (four items) and social support (six items). The response categories are frequency based. Although used exclusively in the public sector, it is anticipated that the scales would apply equally across different sectors and jobs as does the original Karasek measure (JCQ).

The papers reviewed provided limited information on the specific items included in the measures. Extensive norm data exists in the published articles.

4.10 Occupational Stress Indicator (OSI) Sources of Pressure Scale

The Occupational Stress Indicator (Cooper, Sloan and Williams, 1988) was designed to aid organisations in the diagnosis of stressful working conditions. The OSI is based on a model of occupational stress which identifies sources of pressure (experiences in the workplace) as causing stress effects (low job satisfaction, poor mental and physical health) which are moderated by individual differences (coping skills and stress prone personalities). It differs from the other measures reviewed in that the full OSI consists of six questionnaires which attempt to measure four different areas: workplace sources of pressure; individual differences; coping strategies; and stress outcomes. This review, however, is concerned only with the reliability and validity of the sources of pressure scale.

Table 4.6: OSI Sources of Pressure Scale

Name of scale	Highest Alpha	Lowest Alpha	Mean Alpha
Factors Intrinsic to the Job	.76	.63	.69
Managerial Role	.84	.76	.81
Relationships with Others	.74	.87	.79
Career and Achievement	.77	.82	.79
Organisational Structure and Climate	.82	.86	.83
Home-Work Interface	.83	.87	.85

Source: IES, 2000

The OSI sources of pressure scale has a total of 61 items and consists of six sub-scales:

- factors intrinsic to the job (nine items)
- managerial role (11 items)
- relationships with others (ten items)
- career and achievement (nine items)
- organisational structure and climate (11 items)
- home work interface (11 items).

Respondents are asked to respond on a Likert type scale the degree of pressure each item causes them.

The OSI is perhaps the best known measure of its type in the UK and is widely used both in research and commercially within organisations. As a result of its widespread use, many papers were initially identified in the literature (which had used the OSI). However, on reading the full paper, it became clear that only 14 of the original 47 papers were suitable for inclusion in the review. Twenty-one were excluded on the basis that they did not meet the original criteria for inclusion in the review. A further seven papers had to be dropped because they did not provide relevant information (such as response rates or reliability coefficients for the OSI sources of pressure sub-scales) and five papers were excluded because sub-scales had been combined or items dropped from the original measure.

4.10.1 Reliability

Six of the 13 papers included in the review reported on the reliability of the scales. Reliabilities were generally good, although dropped below acceptable levels on two of the six studies for the 'factors intrinsic to the job' sub-scale.

No evidence was found on inter-rater reliability or test re-test reliability or sensitivity.

4.10.2 Validity

Face validity

The measure asks respondents to indicate the degree to which each of the items causes them to feel pressure on a six point Likert type scale.

The scale was developed with largely white collar and managerial workforces, and there is little evidence of attempts to pre-test items on a target population. As a result, the scale might be of less relevance in non-managerial settings.

Content validity

Items for the OSI sources of pressure scales were based on Cooper and Marshall's (1976) model of occupational stress. The 61 items therefore represent the sources of stress described in the model which in turn was based on available stress literature.

Construct validity

Three of the papers reviewed the factor structure of the OSI sources of pressure scale. All three studies reveal serious problems with the proposed structure of the measure (six sub-scales). If an instrument is valid, then the items should assess key aspects of psychosocial hazards in coherent ways. So, if the OSI sources of pressure scale actually measures sources of stress in the way described by the sub-scales (*ie* factors intrinsic to the job, managerial role, relationships with others, career and achievement, organisational structure and climate, and home work interface) then items measuring, for example, career and achievement, should produce similar answers to each other; items measuring home work interface should produce similar answers to each other; and so on. If this is the case, then the instrument's 'structure' conforms to theoretical expectations.

Work by Williams (1996) one of the original authors of the OSI, indicates that the proposed structure (*ie* the six sub-scales) is a poor fit to the data. In other words, similar items such as, for example, those measuring home-work interface do not produce similar answers to each other as would be expected. This is confirmed by two further studies, one by Davis (1996) whose analysis of the 61 items identified four factors as opposed to the six scales proposed by Cooper, Sloan and Williams (1988). Davis suggested that the four areas of work that were being measured were:

- managerial responsibility
- organisation culture
- work demands
- personal demands of work.

The main factor in Davis' analysis (managerial responsibility) accounted for a lot of the variation in the way people responded to the scale indicating that there might be a single underlying explanation for responses. These four factors when considered as scales had better reliability than the original scales proposed by Cooper which leads Davis to conclude that the four scales might be more useful in practical applications of the measure.

The third study which casts doubt on the construct validity of the OSI sources of pressure scale was conducted by Lyne, Barrett, Williams and Coaley (2000) who, like Davies before, found 'no correspondence' between the patterns of responses found in the data and the sub-scales suggested by Cooper and Bramwell (1998). Lyne and colleagues found that statistical analysis suggested the sources of pressure scale in fact was best interpreted as three, or possibly four, sub-scales consisting of:

- workload
- pressures in the role of employee
- pressures of the managerial role
- (lack of support from home).

However, Lyne, Barrett, Williams and Coaley (2000) found that this was not a ideal solution as many of the items (questions) were complex and related to more than one sub-scale, or did not relate to any of the sub-scales when analysed statistically. In the three factor (sub-scale) structure ultimately adopted by Lyne Barrett, Williams and Coaley (2000) they note the similarity of the main workload sub-scale to the 'psychological demand' element of Karasek's model which is reported on elsewhere in this chapter.

Lyne Barrett, Williams and Coaley (2000) conducted further psychometric analyses on the sources of pressure scale in its originally proposed format (as it is currently recommended for use). The results were poor and lead them to conclude…

> *'These results are an emphatic demonstration of the problems with the published OSI sources of pressure score key.'*

Whilst none of the factor analyses reported on here were of particularly high quality when rated against our psychometric criteria, their findings are none the less a cause for concern. Further evidence of the problematic nature of some items in the OSI sources of pressure sub-scale comes from recent work conducted on Dutch samples (Evers, Frese and Cooper, 2000). They identify problems such as items that were too abstract, some that were in the wrong sub-scale and some that were simply not appropriate for most jobs. This led them to develop new items for the OSI sources of pressure scale. Whilst this information should be treated with some caution as it is based on a Dutch translation of the OSI, the authors endorse the use of an English version of their revised OSI scales so it is to be assumed that the same

problems apply to current English versions of the OSI sources of pressure scale.

In summary, this points to considerable problems with the structure of the OSI sources of pressure scale. All three studies which have undertaken factor analyses of the structure do not confirm the six sub-scales proposed by the authors. This is important because it casts doubt on whether the scales are really measuring what they set out to. The separate work by Davies and Lyne suggest three or four sub-scales exist, in both cases one of the sub-scales is concerned wholly with the experience of being a manager (and therefore of limited relevance to the general working population). Two of the original authors of the OSI have separately produced work which reveals problems with the structure of the OSI sources of pressure scale and Cooper (Evers, Frese and Cooper, 2000) has recently proposed the use of alternative items/scales which differ from the published version. Williams has attempted to address these problems through the production of a new measure (Williams and Cooper, 1998 — see also Section 5.5).

Concurrent validity

Given the difficulties with the structure of the OSI sources of pressure scale, other forms of validity are unlikely to be strong. Of the five papers which provided evidence from which the concurrent validity can be assessed, only one (Sparks and Cooper, 1999) consistently demonstrates significant relationships between the six OSI sub-scales and measures of physical and mental health in a way that support the Cooper and Marshall (1975) model of stress. Sutherland and Cooper (1993) found that role factors were related to anxiety only; career achievement and organisational structure and climate were related to job dissatisfaction; and home work interface was related to depression and somatic anxiety. However, it was also associated with significantly better job satisfaction. Cooper, Clarke and Rowbottom (1999) found that some OSI sources of pressure sub-scales predicted better well-being, contrary to expectations. Bradley and Eachus (1995) found that five of the six OSI sources of pressure sub-scales were associated with different measures of harm (job satisfaction, physical and mental health). However, two of the correlations were in the opposite directions to expectation (those for career achievement and factors intrinsic to the job). Bogg and Cooper (1995) found five out of the six OSI sources of pressure scales were associated with measures of harm in the expected way.

Overall this suggests a somewhat mixed picture for concurrent validity. Correlations with measures of harm are nearly always found, but not always in the expected direction. Given the problems around construct validity it is not always clear what some of the sub-scales are tapping.

Predictive validity

No data on predictive validity were found for the OSI sources of pressure scales.

Discriminant validity

On the whole, the majority of papers do not report on discriminant validity. It is possible to infer a differential pattern of relationships for the six sources of pressure scales (indicative of good discriminant validity) in a number of the papers reviewed (Sutherland and Cooper, 1993; Bradley and Eachus, 1995). However, patterns of relationships are not always in the anticipated directory.

Additionally, Evers, Frese and Cooper (2000) report high correlations between sources of pressure sub scales and acknowledge that their proposed shorter version

> *'may make the OSI in total less an indicator of stressful working conditions and more an indicator of personality characteristics and personal well-being.'*

This would pose serious problems for discriminant validity, suggesting that it is a measure of personality characteristics rather than objective work conditions.

4.10.3 Utility

The OSI is perhaps the best known and most widely used measure of workplace stress. It was designed specifically to aid the diagnosis of stress in organisations. The OSI sources of pressure scale, which is the only part of the OSI to be reviewed here, consists of 61 items across six sub-scales.

The OSI has been widely used and researched and as a result it is relatively well understood psychometrically compared to some of the other measures included in this review. However, this research also reveals several issues of concern relating to use of the OSI.

The OSI proposes six sub-scales measuring sources of pressure, however research to date into the structure of the OSI does not support this structure. Data on concurrent validity is also mixed with inconsistent or unexpected relationships being reported (*eg*. high scores on sources of pressure being associated into better well-being). Overall, levels of validity are relatively low for the OSI. As with some of the other measures reviewed in this section the OSI does not use a frequency based response format.

The OSI has extensive normative data and is available commercially.

4.11 Rizzo and House Measures of Role Conflict and Role Ambiguity

Work into role dynamics and their impact on commitment, satisfaction and performance within the workplace dates back to 1964. Kahn first proposed the theory of organisational role dynamics. Rizzo, House and Lirtzman were among the first to tackle the task of developing measures of these potential workplace hazards (Rizzo, House and Lirtzman, 1970). Rizzo and House's work spans the last 30 years and focussed on the relationships between role ambiguity (sometimes referred to as role clarity), role conflict and other theoretically related measures, such as leadership, satisfaction and anxiety. Since they were first presented in a 1970 edition of *Administrative Science Quarterly*, the Rizzo and House Role Ambiguity/Conflict scales have been widely used and appear in many research studies in one form or another.

Rizzo, House and Lirtzman (1970) state that original scales were developed in response to the recognition that…

> *'…The literature indicates that dysfunctional individual and organisational consequences result from the existence of role conflict and role ambiguity in complex organisations. Yet, systematic measurement and empirical testing of these role constructs is lacking.'*

Given their age and widespread use, it is no surprise that the literature search identified 29 papers for possible inclusion in the review. However, on reading the full papers it became apparent that only 13 of the original 29 papers could be considered appropriate for inclusion in the review, the remainder failing to meet the criteria for inclusion in the study (largely due to the fact that they reported one-off adaptations or changes to the original scales).

The original scales presented by Rizzo, House and Lirtzman (1970) consisted of 14 items following analysis of responses to 29 items derived directly from theory and previous research in this area. The two scales were each designed to reflect different aspects of role conflict (eight items) and ambiguity (six items) which are summarised in the box that follows. Respondents were presented with the items in a statement format and asked to indicate on a seven point scale (from very false to very true) the extent to which each statement existed for them in their work.

Role conflict:

- Between the focal person's internal standards or values and the defined role behaviour.
- Between the time, resources, or capabilities of the person and the defined role behaviour.
- Between several roles for the same person which require different or incompatible behaviours, or changes in behaviour as a function of the situation.
- Between the expectation and demands of the organisation through incompatible policies or conflicting requests from others.

Role ambiguity:

- Predictability of the outcomes or responses to one's behaviour.
- The existence or clarity of behavioural requirements which serve to guide behaviour and provide knowledge that behaviour is appropriate.

4.11.1 Reliability

Ten of the reviewed articles contained data on the reliability of the role conflict and role ambiguity scales, covering 20 studies in total. These studies consistently report high levels of reliability for the two scales. Reliabilities for role conflict are above the minimum threshold in 18 out of the 20 studies, and above .80 in half of them. Reliabilities for role ambiguity are equally impressive – above .70 in all but one study, above .80 in 10 of the studies. Summary details are given in Table 4.7 below.

One of the studies reviewed included data on test re-test statistics. The measures were taken several months apart and coefficients were low indicating that the scales had good test re-test sensitivity and are sensitive to change.

4.11.2 Validity

Face validity

Items for this scale were developed from organisational role theory. The original paper describes how 14 items were selected from an original 29, but there is no evidence of attempts to develop or pre-test items on the target population. This is a scale

Table 4.7: Scale reliabilities for role conflict and role ambiguity scales

Name of scale	Highest Alpha	Lowest Alpha	Mean Alpha
Role Conflict	0.87	0.56	0.77
Role Ambiguity	0.89	0.63	0.78

Source: IES, 2000

that was developed in the US over 30 years ago. This does therefore raise some questions about the relevance of the language used in the measure for a UK sample in the working environment of today.

Content validity

This measure is based on theory which is now more that 30 years old. Thinking on psychological well-being in relation to work has undergone radical change and development in that time. More recent papers confirm that ambiguity and conflict are helpful concepts in understanding some of the ways in which work characteristics can have a negative impact on performance and well-being. Inevitably some of these studies also point to the limitations of this approach. The range of items included in the Rizzo and House scales is small in relation to the range of potential psychosocial hazards. Additionally, the measure does not ask how often something is a problem — only how much of a problem it is.

Construct validity

Four of the papers reviewed look in detail at the construct validity of the role conflict and role ambiguity scales. The findings from the papers are mixed, but broadly support the idea of two distinct measures (ambiguity and conflict) as proposed.

Concurrent validity

Seven of the papers reviewed reported on the concurrent validity of this measure. In the main, the expected relationships were found between the role ambiguity and conflict scales and a range of attitudinal measures such as job satisfaction and organisational commitment. The expected relationships were also found between these scales and measures of harm — higher role conflict and ambiguity were associated with greater reports of psychological symptoms. The research also indicates that role conflict and ambiguity are associated in the predicted ways with other potential hazard measures (*eg* organisation practices and leadership behaviour).

Importantly, in most of the research, the patterns of relationships with other measures differ for the two scales, indicating that they are measuring two distinct stressors (as opposed to different aspects of the same thing).

Predictive validity

No predictive validity data were found for these measures.

Discriminant validity

Discriminant validity data were mixed for role conflict and role ambiguity scales. Two studies in particular suggest good evidence of discriminant validity (Kelloway and Barling, 1990; Smith, Tisak and Sneider, 1993). Where as other results point to a lack of discrimination (Schuler, Alday and Brief, 1977). Hall and Spector, 1991 point out that the pattern of responses on these and other measures remains consistent across people in similar and different jobs, raising questions about the extent to which the work environment is the sole cause of the observed relationships.

4.11.3 Utility

These measures of role ambiguity and role conflict are some of the most long established. Even so, recent research suggests that they are still useful in aiding our awareness of the existence of certain workplace stressors.

Two major issues in relation to assessing workplace hazards are that the scales do not use a frequency based response format and that they only cover a limited part of the spectrum of psychosocial workplace hazards.

The scales are brief — eight items measuring conflict, six measuring ambiguity — and easy to complete. Respondents are asked to indicate the extent to which each of the 14 statements is true of their job on a seven point scale ranging from very false to very true. Typical conflict items include:

'I receive an assignment without the manpower to do it.'

Typical ambiguity items include:

'I feel certain about how much authority I have.'

Extensive research evidence points to the scales' reliability. However, there is no recent evidence that the items in the scale remain valid in today's work environments (face validity) and words such as 'manpower' can seem rather dated. The focus of the scale on roles inevitably means that it does not cover all potential areas of hazard (content validity). The structure of the scale (construct validity) appears good and evidence on the relationship between this measure and other theoretically related scales (concurrent validity) was strong. There is no evidence of the ability of this measure to predict whether or not subsequent harm will occur (let alone the extent of such harm) following exposure to the hazards identified. The measures have been widely used across many different occupational settings and seem applicable to different sectors and jobs.

The measures are freely available and well documented in the literature.

5. Information About Other Measures

This section reviews an additional 11 measures for which only limited evidence for the validity, reliability and utility was found.

These are:

- Effort-Reward Imbalance
- NHS Measures
- NIOSH Generic Job Stress Questionnaire
- Occupational Stress Inventory
- Pressure Management Indicator
- Role Hassles Index
- Stress Audits
- Stress Diagnostic Survey
- Stress Incident Record
- The Stress Profile
- Work Environment Scale.

For each measure a brief outline is presented covering, where known, the developers' aims and purposes, development procedure, and the extent and nature of the tool's use. The characteristics of each tool are then itemised, *eg* the number of items and sub-scales, how the test is administered and scored, and the availability of normative data. However, it must be recognised that the quantity and quality of information available varies considerably from measure to measure and hence different information is provided for different measures.

5.1 Effort-Reward Imbalance (Siegrist and Peter, 1994)

5.1.1 Background

What were the aims and purposes of the measure according to the developers?

Simply put, the Effort-Reward Imbalance (ERI) model conceptualises the causes of employee well-being in terms of the relationships between 'costs' or efforts and 'gains' or rewards. Siegrist (1994) describes ERI as having some similarity with Karasek's job strains model, but also points to important differences:

> *'First, its focus is not on job task content but on the reward structure of work.'* (Siegrist and Peter, 1994, p.131)

According to Siegrist the imbalance between high effort spent at work and low reward is a particularly damaging to well-being as this:

' … violates core expectations about reciprocity and adequate exchange.' (Siegrist, 1996, p. 28)

Siegrist goes on to distinguish between immediate rewards (*eg* economic rewards, socio-emotional feedback) and the longer term expectations of rewards termed 'status control'.

A further important feature of the ERI model is that it seeks to incorporate personal coping styles into the model.

How was it developed?

ERI approaches were primarily developed in studies of cardiovascular health.

However, there does not appear to be a standard instrument. No measures of reliability are reported on in any of the studies. With all studies it was not possible to identify a consistent form of measurement.

Evidence about its use

Several studies are available which report on using this approach, mostly in relation to physical health outcomes.

Number of studies/reports used to examine this measure

Eleven papers were identified which to a greater or lesser extent describe using this approach.

5.1.2 Utility and other information about the measure itself

Not applicable.

5.2 NHS Measures (Haynes, Wall, Bolden, Stride and Rick, 1999)

5.2.1 Background

What were the aims and purposes of the measure according to the developers?

The aim was to develop short scales for use in the NHS, with good face validity, clear factor structure and high internal reliabilities. The scales were designed to measure autonomy/control, feedback on work performance, influence over decisions, leader support, role clarity, role conflict, peer support, and work demands. A final construct of professional compromise was identified through pilot interviews.

How was it developed?

The constructs were taken from leading theoretical frameworks. A qualitative development phase 'to determine and develop the applicability of a large number of potential items' followed (Haynes, Wall, Bolden, Stride and Rick, 1999, p. 260). Pilot questionnaire administered to 825 healthcare employees and Exploratory Factor Analysis used to identify to reduce factors.

Evidence about its use

Used in large NHS project of around 11,600 participants.

Number of studies/reports used to examine this measure

Reports all likely to be linked to same project.

5.2.2 Utility and other information about the measure itself

Numbers of items and sub-scales

Forty-three items, nine sub-scales.

Sample items

See Haynes, Wall, Bolden, Stride and Rick (1999, pp. 273-275) for full item list.

Availability and costs

The measure is freely available (see above).

Administration and scoring

Self-completion, simple scoring with some reverse scored items. Uses a Likert response format.

Applicability to sectors, jobs, *etc.*

All scales widely applicable except for the NHS specific Professional Compromise scale.

Availability of normative data

Normative data based on over 8,000 respondents is provided in Haynes, Wall, Bolden, Stride and Rick (1999).

5.3 NIOSH Generic Job Stress Questionnaire (Hurrell and McLaney, 1988)

5.3.1 Background

What were the aims and purposes of the measure according to the developers?

The main focus is on 13 job stressors as well as measures of distress and modifiers. Sections have been developed to be used in a modular fashion.

How was it developed?

Specific stressor, distress, and modifier variable constructs were selected for inclusion on the basis of a content analysis of the job stress literature. The scales selected to measure these constructs were adapted from scales with known reliability and validity.

Evidence about its use

Various scales have been used widely, though few studies published including all measures. It has also been translated into several languages (including Japanese: Iwata, Kawakami, Haratani, Murata and Araki, 1999).

Number of studies/reports used to examine this measure

Difficult to determine as the instrument is modular, various studies have used different elements of it.

5.4 Occupational Stress Inventory (Osipow and Spokane, 1987)

5.4.1 Background

What were the aims and purposes of the measure according to the developers?

A generic measure of stress, strain and coping resources phrased in an occupational context.

Evidence about its use

One investigation which used this measure was found (Fogarty, Machin, Albion, Sutherland, Lalor and Revitt, 1999). It comprised three studies conducted in Australia on 153 employees from a variety of sectors, 98 healthcare workers and 107 military personnel. Limited internal reliability and concurrent validity information was available from this report. Additional normative information is available from the manual (Osipow and Spokane, 1987).

Number of studies/reports used to examine this measure

Two.

5.4.2 Utility and other information about the measure itself

Numbers of items and sub-scales

The Occupational Stress Inventory has 147 items in three questionnaires measuring occupational stress, strain and coping resources. The Occupational Roles Questionnaire assesses stress, with six sub-scales measuring role overload, role insufficiency, role ambiguity, role boundary, responsibility, and physical environment. The Personal Strain Questionnaire assesses occupational strain with four sub-scales measuring vocational strain, psychological strain, interpersonal strain, and physical strain. The Personal Resources Questionnaire assesses coping skills with four sub-scales measuring recreation, self-care, social support and rational/cognitive coping resources.

Administration and scoring

All items are self-report format. The response scale is a five point Likert type, ranging from 'rarely or never true' to 'true most of the time'. The 14 scales can be used separately or summed to produce measures of stress, strain and coping.

Applicability to sectors, jobs, *etc.*

Generally applicable, most relevant to white collar workers in medium to large organisations.

Availability of normative data

Internal consistency for the 14 sub-scales ranges between 0.71 to 0.94 (Osipow and Spokane, 1987). Other normative information is available in the manual. When summed into the three main indices, internal consistencies are higher (*eg* Fogarty, Machin, Albion, Sutherland, Lalor and Revitt [1999] reported 0.89, 0.86 and 0.93 for stress, coping and strain).

5.5 Pressure Management Indicator (Williams and Cooper, 1998)

5.5.1 Background

What were the aims and purposes of the measure according to the developers?

The Pressure Management Indicator is a recent evolution of the OSI (see Section 4.1), designed to overcome the limitations of the original instrument.

How was it developed?

Firstly the psychometric properties of the OSI were developed through an iterative process of exploratory factor analysis, item analysis and confirmatory factor analysis. Secondly, extra items were added 'designed to strengthen the weaker scales and produce additional scales' (Williams and Cooper, 1998). Measures of organisational commitment, job security and decision latitude were added.

Evidence about its use

Only limited data exists as yet. Some preliminary findings using the PMI were published in Williams and Cooper (1998). However, as the authors note:

'As the PMI is a new instrument it is not yet possible to provide a detailed list of research publications.'

Number of studies/reports used to examine this measure

One.

5.5.2 Utility and other information about the measure itself

Numbers of items and sub-scales

Ninety items arranged into 22 sub-scales. Job satisfaction is measured as satisfaction from the job itself and from the organisational climate. Mental and physical health sub-scales are: state of mind (anxiety-depression), resilience, confidence level, physical symptoms and energy levels. Pressure is measured on eight factors: workload, relationships, home-work balance, managerial role, personal responsibility, daily hassles, recognition and organisational climate. Individual differences are drive and impatience (reflecting type A behaviour) and personal influence and control. Coping scales reflect the use of problem focused coping and life-work balance. Social support was identified as a 'separate construct'. All scales except Daily Hassles (0.64) have internal consistencies above 0.70.

Sample items

Largely similar to the OSI.

Administration and scoring

Largely similar to the OSI.

Applicability to sectors, jobs, *etc.*

Generally applicable.

5.6 Role Hassles Index (Zohar, 1997)

5.6.1 Background

What were the aims and purposes of the measure according to the developers?

The Role Hassles Index was 'designed to reflect episodes of role conflict, ambiguity and overload' (Zohar, 1997).

How was it developed?

Items and definitions of stressors were derived from major questionnaires and literature reviews. Five participants then described 96 events or hassles associated with each stressor. These were reduced to 35 events following group discussion. A final list of 20 events comprised the RHI. Exploratory factor analysis was used to confirm structure.

Evidence about its use

The first study comprised 161 hotel employees in Canada. Internal consistencies were 0.80, 0.71 and 0.82 for conflict, ambiguity and overload respective indicating good reliability levels for the three scales. However, results also suggested low concurrent validity with mental health measuring and medium association with a Perceived Stress Scale (Cohen, Kamarck and Mermelstein, 1983).

Number of studies/reports used to examine this measure

Two.

5.6.2 Utility and other information about the measure itself

Numbers of items and sub-scales

Twenty items, three sub-scales (conflict, load and ambiguity hassles).

Sample items

> *'Had an argument or confrontation over differing views'; 'Felt under time pressure, had difficulty due to insufficient time'; 'Had concerns about how to solve a problem'.*

Administration and scoring

Self completion. Subjects rate events experienced over previous two weeks in terms of how disruptive (physically or emotionally). They were using a three point scale (slightly, quite, very disruptive). Scoring 'consisted of adding up the severity ratings of the reported items in each scale' and 'dividing by the total sum of that scale'.

Applicability to sectors, jobs, *etc.*

Generally applicable to those working within organisations.

Availability of norm data

None, apart from the published reports.

5.7 Stress Audits (Sutherland and Davidson, 1993; Sutherland and Cooper, 1996; Lancaster, Pilkington and Graveling, 1999)

What are the aims and purposes of the measure according to the developers?

Stress audit is a generic term which describes a number of broadly similar approaches. The purpose of the stress audit method was to: '1. Identify the potential sources of stress (*ie* the stressors); 2. Assess which of the sources of stress have the greatest negative impact; 3. Identify which, if any, of the individuals and or groups of workers…have particular stress-related problems.' (Sutherland and Cooper, 1996, p. 27-28).

How was it developed?

The term 'Stress Audit' does not refer to a specific instrument used rather it refers to a process. Stressor items were 'generated from interviews conducted in the qualitative phase of the study' (Sutherland and Davidson, 1993, p. 276). This process was repeated in Sutherland and Cooper (1996) using 50 offshore workers. Thus different questionnaires were developed for each study.

The Lancaster, Pilkington and Graveling approach describes three phases: this first is the identification of hazards via semi-structured interviews. Stage two involves the detailed investigation of priority areas and stage three the evaluation of any intervention activity.

Evidence about how much it is/has been used and in what contexts

There would seem to be evidence that this approach has been used extensively, but the precise instruments used have varied according to the employee group.

Number of studies/reports used to examine this measure

Two empirical studies, a third paper (Cooper and Cartwright, 1997) contained no empirical information and was not used. An IOM report part funded by HSE, provides detailed case study evidence in the Organisational Stress Health Audit (OSHA) (see Lancaster, Pilkington and Graveling, 1999 for details).

Numbers of items and sub-scales

The number of items varied in the two studies. There were 36 stressor items in Sutherland and Davidson and over 75 items described in Sutherland and Cooper.

Sample items

Factor analysis identified eight main factors for oil rig workers (career prospects, safety, home-work interface, under-stimulation, physical conditions, unpredictability, living conditions, physical climate) plus four additional factors (organisation structure and climate, physical well-being, workload, air transportation). Eight factors were found by Sutherland and Davidson (ambiguity, overload, manpower problems, culture and problems, homework interface, role insecurity, boundary relationship, new technology).

Administration and scoring

Administration of self completion questionnaires by mailshot. Factor scores derived by adding related items. Items were scored on a five point rating scale (Sutherland and Davidson used 'no stress' to 'high stress', Sutherland and Cooper used 'no pressure' to 'high pressure').

Applicability to sectors, jobs, *etc.*

The stress audit methodology described is widely applicable, though the instruments derived are sample specific.

Availability of normative data

Only limited normative data would seem to be available for each work sector.

5.8 Stress Diagnostic Survey (Ivancevich and Matteson, 1980)

5.8.1 Background

What were the aims and purposes of the measure according to the developers?

It was designed to help individual employees identify specific areas of high stress at work.

How was it developed?

Developed from exploratory factor analysis of work stressors from a range of occupational groups including business executives, health care workers, and graduate managerial and engineering students.

5.8.2 Utility and other information about the measure itself

Numbers of items and sub-scales

Work version consists of 80 statements, resulting in 15 sub-scales of work stressors.

Administration and scoring

Self completion. Seven point Likert type scale, anchored with a end and mid-points (never, sometimes or always a source of stress).

5.9 Stress Incident Record (Newton and Keenan, 1985; Narayanan, Menon and Spector, 1999)

5.9.1 Background

What were the aims and purposes of the measure according to the developers?

The Stress Incident Record is not based on scale measures. Rather its aim is to focus on specific stressful incidents rather than typical circumstances at work, using an open-ended method.

How was it developed?

As a qualitative instrument, the development phase of the instrument was not extensive.

Evidence about its use

Few specific reports focusing exclusively on this method were found. However, the general method is likely to have been more widely used, perhaps in pilot studies.

Number of studies/reports used to examine this measure

Two.

5.9.2 Utility and other information about the measure itself

Not applicable.

Administration and scoring

In Narayanan, Menon and Spector (1999), participants 'were asked to describe the most stressful incident that occurred at work over the past one month that made you feel anxious, annoyed, upset or frustrated, or aroused your feelings in any other way' (p.66). Participants were also asked to rate how stressful this event was on a 4-point scale ranging from 'not very' to 'very much'.

Applicability to sectors, jobs, *etc.*

Widely applicable.

Availability of normative data

Not applicable.

5.10 The Stress Profile (Setterlind and Larsson, 1995)

5.10.1 Background

What were the aims and purposes of the measure according to the developers?

> *'The Stress Profile is a psychosocial instrument for measuring stress in life in general and at work at the levels of the individual, the group and the organisation.'* (Setterlind and Larsson, 1995, p.85)

How was it developed?

An initial pool of 300 questions was tested on 500 subjects. 'On the basis of the statistical analysis all unreliable questions were deleted. The remaining 250 were subjected to factor analysis' (Setterlind and Larsson, 1995, p. 87). The reduced Stress Profile was tested on a new group of 400 subjects and these results cross-checked for validity against the first sample, reducing the profile to 224 items.

Number of studies/reports used to examine this measure

One.

5.10.2 Utility and other information about the measure itself

Numbers of items and sub-scales

224 items, including 20 background variables and ten criteria. There are 16 main fields divided into 60 subsidiary fields. These main fields include external causes of stress (psychosocial work environment, work content, workload and control, leadership climate; physical work environment, family relationships, major life events, daily hassles/satisfactions); internal causes of stress (self-perception, sense of coherence); coping with stress (problem-focused, emotion-focused, type A behaviour, lifestyle); stress reactions (physical, emotional, cognitive, behavioural burnout).

Administration and scoring

Scored using a 'specially designed computer program'.

Applicability to sectors, jobs, *etc.*

Likely to be applicable widely.

Availability of normative data

Normative information based on 4,000 cases available.

5.10.3 Number of studies/reports

At least two other instruments exist called the Stress Profile — one by Derogatis (1984), the other by Wheatley (1990) — were identified during the search procedure. However, these two were excluded from the review as the instrument and scales were not about work and there is little evaluative literature.

5.11 Work Environment Scale (Moos, 1994)

5.11.1 Background

What were the aims and purposes of the measure according to the developers?

Not really designed to measure job stress but developed to assess the general work climate.

Evidence about its use

It has been used widely, with an emphasis within treatment and care agencies.

Number of studies/reports used to examine this measure

Two.

6. Conclusions and Recommendations

Thus far, this review has considered evidence for the reliability and validity of a range of psychosocial hazard measures. It has not yet summarised this evidence nor considered its implications for research and practice. This chapter concludes the report by providing a brief overview of the evidence presented in detail earlier and considers what this evidence means for both future research and future practice. Before this is done, the objectives of the review are restated, and a description is provided of the kinds of evidence and measures that were found.

6.1 Review objectives and method

Organisations measure psychosocial hazards (or 'stressors') for numerous reasons and many tools and instruments that purport to measure hazards are available. However, little systematic information about the quality of these measures is available. When considering the quality of psychometric measures it is usual to consider their reliability (*ie* consistency or accuracy) and their validity (*ie* meaningfulness or relevance).

The aims of the review were therefore to:

- identify the methods or measures currently available to assess workplace psychosocial hazards
- look at each of the measures identified and assess their reliability and validity
- assess the utility of different measures.

In order to meet these aims the review undertook a number of tasks. First, extensive literature reviews were undertaken in order to identify relevant measures and sources of information about their reliability and validity. Second, criteria for assessing reliability and validity were developed. Third, a team of reviewers applied these criteria to the available information about the measures. Last, evidence about the reliability and validity of five main measures and a number of other measures was collated.

6.2 What evidence was available?

A surprising finding of this review, given the many thousands of research papers on occupational stress produced over the last thirty or more years, is the general lack of serious (replicated) studies examining the psychometric properties of measures of psychosocial hazards. While there were many studies which used these measures, they did not often include information which could be used in this review. This was for two main reasons.

First, there was inconsistent reporting of reliability and validity data in many articles. Although measures were widely used, reliabilities for the scales were not always given and often, quite basic information about response rates to surveys and correlations between scales was lacking.

Second, there was also inconsistent use of measures. In many cases, original measures had been adapted, items added or deleted or the response format changed. Whilst this had often been done for very good research reasons (*eg* to try and improve reliability in a specific research setting), such applications are not helpful for practitioners seeking to use 'off-the-shelf' instruments. This was also particularly true for the measures of demand and control. Some reliability and validity data that were identified, therefore, related to one off or adapted measures and could not be incorporated into the review.

Although many papers refer to and draw on the psychosocial hazard measures covered in this review, the number of papers which, when examined closely, actually provide psychometric information on the properties of the relevant measures is surprisingly small.

There are three important implications of this finding:

- The number of measures about which there is sufficient information to provide a detailed review of psychometric properties is small — this study identified only five.
- There are many other measures for which only very limited information was available — evidence about these measures has therefore been reviewed in less detail.
- In general, the quantity and quality of evidence relating to the reliability and validity of hazard measures is limited. This means that in some cases only tentative conclusions can be drawn about reliability and validity. It also means that for many of the measures currently in use there is simply no significant body of evidence about their reliability and validity.

6.3 What measures are available?

A striking finding of this review was the lack of variety in the type of psychosocial hazard measures that have been developed and used. Extensive searches of the literature and discussions with professional bodies revealed that by far the most common type of hazard measurement was the self-report questionnaire. In addition, nearly all of these were designed primarily for research and not as organisational tools.

Typically, in such measures, items describing a hazard (*eg* workload) are presented, and respondents are asked to rate the extent to which they agree or disagree with a statement describing the hazard. There is also relatively little variety in the content of items or the particular type of response scale used. While there have been attempts to develop alternative methods this is not currently well advanced. The implications of this somewhat narrow approach will be considered later.

Given the broad range of work conditions that could potentially be psychosocial hazards it is not surprising that a large number of hazards has been measured. Three of the main measures reviewed, the Job Stress Survey, the Job Diagnostic Survey and the OSI Sources of Pressure scale all attempt to gather information about many different hazards. However, two of the other main measures reviewed, Karasek's Demand and Control measures and the Rizzo and House measures of role conflict and role ambiguity, each focus on just two types of hazards.

In general, although a wide range of hazards have been measured it is not always clear why they have been chosen or if important psychosocial hazards remain unmeasured.

While some of the very generic measures of hazards could apply to any jobs sometimes appear to apply more to white-collar jobs than manual work. While there are measures of more specific hazards, these hazards do not necessarily apply to specific kinds of jobs or occupations and in the measures reviewed here there appear to be relatively few job-specific measures of psychosocial hazards even though it may be the case that jobs do contain relatively unique kinds of hazards.

In general, measures tend to assess broadly the same kinds of hazards and do so using very similar measurement techniques. This means that many possible approaches have never been tried or tested and that, inevitably, this review reflects the somewhat limited approaches adopted this far.

6.4 Evidence for reliability

Reliability is connected with the consistency of the measurement. There are a number of different forms of reliability and each of these was considered when examining the hazard measures.

- **Internal consistency reliability**: this refers to the extent to which items within a scale tend to be answered in similar ways and can therefore be considered to be measuring the same sort of hazard. In general the internal consistency reliability was reasonably good across all the measures reviewed. However, it should be noted that internal consistency is relatively easy to achieve In addition, there is something of a bias in the available evidence in that studies which use scales with low internal consistency are unlikely to get published. Hence, the available evidence on this form of reliability will almost inevitably be positive.
- **Test-retest reliability**: this refers to whether measures taken across time remain consistent where and when we would expect them to do so. In general, little evidence was available about test-retest reliability. This is potentially an important form of reliability for hazard measurement as it is helpful to know whether the measures do remain consistent over time where we have no reason to expect them to change.
- **Test-retest sensitivity**: this refers to whether measures taken across time *change* where and when we would expect them to do so. In general, little evidence was available about test-retest sensitivity. This type of reliability is important for hazard measurement as we need to know if the hazard measurement is sensitive enough to pick up changes in psychosocial hazards.
- **Inter-rater reliability**: this indicator of reliability shows the extent to which different people using the same measure to rate the same thing (*eg* a job or task) make similar ratings. In general little evidence was available. However, it is not clear that for most of the measures reviewed here inter-rater reliability is an important form of reliability as most of the measures reviewed are intended to assess individual perceptions.

Some conclusions about reliability

Reasonable evidence of reliability is available for only one of the four kinds or reliability discussed here, internal consistency. Almost no evidence was found for the three other kinds of reliability. This is particularly relevant in the case of test-retest reliability and test-retest sensitivity: Whether psychosocial hazard measures are appropriately consistent or sensitive over time are vital aspects of their reliability about which little is currently known.

6.5 Evidence for validity

While reliability is concerned with the consistency and performance of the measurement, validity asks about the extent to which scales accurately measure what we think they do. In general, more evidence was available for each of the forms of validity than was the case for each form of reliability.

Some aspects of validity are not connected with how the hazard measure performs in practice but rather with the underlying theory or explanation about what the hazard is, how it works, and why it is being measured in the way that it is. As discussed earlier, many forms of validity and, in particular, content and construct validity, are very seriously compromised by the limited and weak theory which underlies some of the hazard measures reviewed.

- **Face validity**: this is where the measure appears to measure what it is supposed to measure to a non-expert. For example, does a measure of workload look like it is measuring workload? In general, face validity appears to be reasonable across the measures as some were developed from interviews or observations hence increasing the likelihood that they are meaningful to non-expert respondents.
- **Content validity**: this is where the measure appears to measure what it is supposed to measure to an expert. As indicated at the start of this section, with limited theory about how the hazard operates, it is difficult to obtain high levels of content validity. Some measures are based on theory and have reasonable content validity in this sense. However, it is clear that almost all of the hazard measures have low content validity in at least one respect — that it they tend to ask about the *extent* to which something is a problem or whether or not a respondent *agrees or disagrees* that something is a problem rather than asking how often or how frequently the problem occurs. In addition, it appears to be the case that most hazard measures have questionable content validity in that it is not clear whether the items used really capture the full range of phenomena that are subsumed under the hazard (*eg* control, workload), or even if they capture the full range of hazards.
- **Construct validity**: a major element of the statistical assessment of construct validity refers to the extent to which a measure behaves as one would expect it to in terms of its structure. A reasonable amount of evidence was available on this form of validity. In general, across all the measures, construct validity is moderate with some measures showing reasonable and others quite poor levels. Other aspects of construct validity are concurrent, predictive and discriminant validity.
- **Concurrent validity**: this refers to the idea that we would expect the hazard measure to be related to other theoretically-

related measures taken at the same time. For example, we might expect that a measure of workload would relate to a measure of fatigue. A reasonable quantity of evidence was available and this indicated that on the whole, the measures showed moderate to good levels of construct validity. However, it should also be noted that it was often the case that hazard measures were related to *many* other variables which can not only suggest good concurrent validity, but also weak divergent validity (see below).

- **Predictive validity**: of all the kinds of reliability and validity thus far discussed, predictive validity would appear to be the most important feature of hazard measures as it refers to whether a measure taken at one point in time predicts theoretically related and important outcomes at some point in the future. In other words, is there evidence that the hazard measures reviewed here actually *predict* future levels of, say, harms such as illness? One of the most significant findings of this review is that there is *very little evidence* about the predictive validity of psychosocial hazard measures. This means that in general we simply do not know whether these measures are valid tools measuring hazards which predict harms.
- **Discriminant validity**: this refers to whether the measure is *not* related to theoretically unrelated variables. As mentioned above, concurrent validity for many of these measures is good in that they are related to theoretically related measures. However, there is also evidence that these measures are also related to other measures to which they are not theoretically related. If measures are related to things they should not be related to, this gives reasonable grounds to question their discriminant validity.

Some conclusions about validity

There is reasonable evidence on which we can make judgements about most forms of validity. What this evidence suggests is that, taking all the measures as a set, the validity of the measures reviewed is at best moderate. Of course, specific measures vary in their level of validity. Much of this limited validity can be traced back to limited attention paid to the theoretical meaning of the measure. Particular problems are found in regard to content validity as many measures show a limited scope of measurement and adopt response formats that may not make sense theoretically.

Most striking however, is the limited evidence for predictive validity. For all but one of the measures reviewed here (which is the Whitehall II measure of demands and control), we simply do not know if the hazard measure predicts important outcomes.

6.6 The utility of hazard measures

Given the similar nature and format of most hazard measures discussed above relatively similar points can be made about their utility. First, they can be administered by anyone and no special training is explicitly required. Second, they are all reasonably easy to complete though some, particularly the generic stressor measures, contain a larger number of items. Third, there are issues around the interpretation of these hazard measures which does somewhat diminish their utility.

Generally speaking, because of issues connected with their limited reliability and in particular validity it is very difficult know what a score on any of these measures actually *means* and therefore what could and should be done in response to it. For example, what, precisely, does a particular score for an individual employee or group of employees on a measure of control tell us? How do we know what we should do about it, if anything, and why?

As the rationale for this review suggests, we need to know about the reliability and validity of available hazard measures in order to assess their utility. If evidence is absent or indicates limited validity then this implies that the utility of these measures will likewise be constrained.

6.7 Recommendations

The main aim of this review was to assess the evidence for the reliability and validity of a range of psychosocial hazard measures. While it is recognised that, in practice, these measures are probably rarely used on their own but supplemented with other forms of investigation and assessment, it remains vital that the measures that are used have reasonable reliability and validity.

The main findings were that:

- compared to the number of papers published on stress and which use measures of psychosocial hazards surprisingly little relevant evidence was found
- there is limited variety in the type of hazards that are measured or the techniques used
- a substantial amount of evidence was available for only one form of reliability, internal consistency, which was reasonably good
- more evidence was available for most types of validity and this indicated mixed levels of validity. There was, however, almost limited evidence for predictive validity.

Broadly speaking, there was relatively little sound evidence about the reliability and validity of these measures. However, what this evidence strongly suggested was that the quality of these measures was limited. This also means that their utility is also likely to be quite limited. These weaknesses have now for the first time been systematically identified. Some of the steps which can be taken to improve such measures of psychosocial hazards are now considered. These recommendations are not comprehensive but focus on those that seem most important and urgent. First the implications for practice and then the implications for research are considered though it is recognised that these areas are interrelated.

6.7.1 Recommendations for practice

On the basis of currently available evidence it is not possible to recommend the use of any of these measures for assessing psychosocial hazards, nor is it possible to identify one measure that is clearly superior to others.

There is a sense that currently these measures are used simply because they are exist and readily available alternatives do not. However, it is not possible to simply stop assessing psychosocial hazards until the required research into existing measures and the development of alternative measures is complete.

The first recommendation is therefore a serious reappraisal of what these measures are actually being used for. Why measure hazards? What is the purpose of risk assessment for psychosocial hazards and what kinds of tools would help with such assessments? What will be done with information which is gathered in this way? How do these assessments fit with other health and safety and human resource policies and practices? Unless issues such as these are further clarified it will not be possible to devise focused and meaningful assessments.

The second recommendation, which follows from the first, is that organisations should be prepared to consider developing their own measures which should be:

- focused on particular organisations and jobs or roles
- more specific and perhaps shorter
- based on local knowledge and understanding of the context
- informed by best practice (*eg*, frequency based response formats)
- incorporated into some form of risk management framework if possible.

Third, it is recommended that organisations should continue to develop other ways of assessing hazards in addition to self-report questionnaires such as:

- observations
- task analysis
- job descriptions
- reports of harms and what these may tell us about hazards.

In general, given the available evidence and what it suggests about hazard measures, organisations need to be much more *proactive* in devising and thinking through their hazard assessment as off-the-shelf measures are likely to have limited utility. A more proactive approach should also help to ensure that local knowledge about specific tasks, jobs, and psychosocial hazards is fully incorporated into the assessment process.

6.7.2 Recommendations for research

As identified in many places in this review there are significant gaps in our knowledge of existing measures and, in particular, a profound absence of knowledge about the predictive validity of most hazard measures. So the first recommendation is that more fundamental validation research is undertaken into some existing measures. There seems little point in doing this with all or most existing measures as some are already known to have limited validity in other respects. However, those measures which have reasonable content validity should be explored further.

However, more important than working with existing measures, are attempts to develop and test new approaches.

The second recommendation is to re-examine theory in the area of psychosocial hazards in order to consider more carefully what we are measuring, and why and how we are measuring it. Some existing theories within the stress field could be helpful. More promising are other theoretical approaches which really try to unpack how particular kinds of work events may lead to emotional and health reactions. If measurement is not based on sound theory then the problems with content and construct validity described above will simply not go away.

A third recommendation is that new and innovative types of measures and methods are developed and tested. There are numerous reasons, described earlier, to suppose that the standard technique of hazard measurement (of presenting an item about a hazard with an agree-disagree response format) has low utility. Hence these new ways of measuring hazards need to be based on theory and also consider item content, response formats, observational methods, checklists, and so on. It may be that work on other forms of hazard assessment could inform the

development of such measures. The testing of these new forms of hazard measure needs to take place in diverse organisational contexts in order to maximise reliability, validity, and utility.

A fourth and urgent recommendation is to examine the measures of harm that are currently used in much the same way as this review has considered measures of hazards. Without reliable and valid measures of harm it is not possible to develop reliable and valid measures of hazards as the sole purpose of assessing hazards is because they are thought to cause harms. If we cannot assess harms in a valid way, we cannot assess the validity of hazard measures.

Last, hazards are not measured in isolation, but as part of a larger process, such as risk management. The fifth recommendation for research is that more attention is paid to these processes so that information gathered about hazards and harms can be better integrated, and the actions taken as a consequence of gathering such information can then be evaluated. While the reliability and validity of hazard measures is essential, their ultimate utility can only be assessed in the context of the processes in which the measures are used.

Appendix I: Psychometric Criteria for Assessing Psychosocial Hazard Measures

As discussed in the main body of the report, all instruments need reliability and validity. That is, all instruments should be capable of producing:

- consistent results free of error (reliability); and
- results that accurately assess what the instrument claims it measures (validity).

This appendix reproduces the materials distributed to the project team which were used to ensure a consistent understanding and reporting of different types of reliability and validity.

The process of statistical assessment of instruments designed to measure psychological variables is often referred to as psychometric assessment or 'psychometrics' (mental measurement) for short (see, for example, Nunnally [1978] for a good overview of psychometric theory and assessment; and Oppenheim [1992], for a good introduction to reliability and validity in the context of questionnaire design).

Ideally, instruments should be subject to several validity studies, preferably including validity studies conducted by research teams that consist of different researchers from different institutions to those that developed the instrument. This is to help ensure the instrument still performs adequately even when different procedures are used (*eg* questionnaire administration and return procedures) (Cook and Campbell, 1976). However, it is rare for independent validation of instruments to be reported in peer reviewed scientific journals. Nevertheless, other studies using an instrument will typically report data on reliability, correlations and associations with theoretically relevant variables from which validity can be inferred. Sometimes such studies also report factor analyses of the scales. Therefore, for widely used instruments, it is possible to develop an overall summative assessment across several studies as has been done for the main measures reviewed in this report.

In the main, measurement of psychosocial hazards is confined to standardised checklists or in which respondents or, very occasionally, independent raters are asked to indicate their choice

of answer to a number of standard questions from a closed list (*eg* rate on a five point scale). Each choice in this list of answers is usually then given a number (*eg* on a frequency scale, 'never' = 1, 'sometimes' = 2 *etc.*). Such data are usually treated as interval level data, but are more correctly viewed as scalar (*ie* falling between interval and ordinal data). This is the case for the standardised instruments reviewed in this report. Qualitative or non-standard methods cannot usually be assessed for reliability or validity in the ways described below, although this is sometime possible (*cf* Daniels, de Chernatony and Johnson , 1995). Nevertheless, where possible, researchers should make every effort to report what psychometric or other evidence they have for the validity and reliability of the methods used.

The first two sections of the appendix give general descriptions of types of reliability and validity. The third section describes the criteria which were used to assess each study. Finally, the criteria used to summarise evidence for each measure from across different studies is presented.

A1.1 Reliability

There are two ways of assessing reliability for self-report instruments: (i) internal consistency reliability is essential for any instrument; (ii) test-retest reliability, which may or may not be appropriate in any given instance. For instruments completed by external raters, a third form of reliability, inter-rater is appropriate.

A1.1.1 Internal consistency

Internal consistency is often assessed in two ways, using Cronbach's alpha or using split half-reliability.

Cronbach's alpha is essentially the average multiple correlation between each item in a scale and all other items in the scale. The Kuder-Richardson coefficient is a special case of Cronbach's alpha for scales consisting of dichotomous variables.

Split-half reliability is the correlation between a (random) set of half the items in a test and the other half of the items. Often, split-half reliability will be corrected using the Spearman-Brown formula: which adjusts for splitting the original full scale in half.

For self-report instruments, scales or subscales of a test must have alphas, split-half reliabilities or Spearman-Brown corrected reliabilities >.70 for acceptable reliability.

This cut-off of 0.70 is the one usually accepted (Nunnally, 1978). However, reliabilities of >.60 but <.70 may be described as marginal. If the majority of scales in a multi-dimensional instrument are >.70, but one or two <.70, then it may be judged

that, on balance, the instrument shows acceptable reliability, with a caveat for those scales with reliability <.70.

A1.1.2 Test-retest reliability

Test-retest reliability is not essential, but desirable. Test-retest reliability is the correlation of a scale with itself measured at some point in the future. Where work environments can be expected to be stable, then test-retest reliability should exceed 0.70. Like internal consistency, a test-retest reliability of <.70, >.60 may be described as marginal. However, there are other exceptions to this rule.

Psychosocial hazard measures should be sensitive to changes in work environments, so where work environments are expected to change, such as after job redesign, test-retest reliability might be low. Therefore, we also sought to assess test-retest *sensitivity* where appropriate.

Similarly, where hazards are rated on response scales over a short time period (say the past week), but the interval between measurements is longer than that period (say three months), test-retest reliability may also be low.

A1.1.3 Inter-rater reliability

Inter-rater reliability is essentially the correlation between two or more independent raters on a response scale. There are several ways of testing for inter-rater reliability, but in all cases, the coefficient of inter-rater reliability should be > 0.70 for acceptable reliability. Again, reliabilities <.70, >.60 may be described as marginal.

Where the data are categorical, rather than interval, ordinal or scalar, then Cohen's kappa (2 raters) or Fleiss' extension to Cohen's kappa (> 2 raters) should be used (see Hays, 1988).

Where there are two raters, appropriate correlation coefficients should be used where the data are interval, scalar (both Pearson's r), ordinal (*eg* Spearman's rho, Kendall's coefficient of concordance) or dichotomous (both Pearson's and Spearman's are suitable here — see *eg* Hays, 1988).

Where there are more than two raters, inter-rater reliability is mostly confined to interval or scalar data. Here, there are several choices:

- Variants of the intra-class correlation (ICCs, see for example Shrout and Fleiss, 1979), which returns one value for each scale assessed. However, ICCs can produce artefactually low coefficients (James, Demaree and Wolf, 1984), so should be treated with some caution.

- r_{wg} (James, Demaree and Wolf, 1984, 1993), which returns an index of agreement across raters for each person on each scale being rated. An average r_{wg} may be reported across all persons rated.

A1.2 Validity

There are essentially three main forms of validity: *face*, *content* and *construct*. (NB: Different texts present slightly different classification and sub-classifications of terms for validity, but all include in general the forms of validity outlined here, *cf* Oppenheim, 1992; Spector, 2000).

A1.2.1 Face validity

Face validity is where an instrument looks like it measures what it should to a non-expert. This is most often assured by developing items from interviews or other qualitative methods. In addition, or alternatively, there may be extensive pre-testing of the instrument with members of the target population. Face validity is a useful characteristic of psychosocial hazard measures — response rates might be higher for instruments that look like they might produce accurate results that might help bring about improvements in working conditions (cf. Oppenheim, 1992).

A1.2.2 Content validity

Content validity is where an instrument looks like it measures what it should to an expert. Here, the instrument should cover the full range of phenomena subsumed under the theoretical definition of that phenomenon (*eg* a measure of work control should cover participation, control over work schedules, work methods, work objectives, rather than just say participation). Further, content validity can include judgements on the adequacy of response formats and item wording. Others have suggested that items with frequency based anchors over specified time periods are likely to have better content validity, as this is likely to minimise cognitive-affective processing (Frese and Zapf, 1988). Others parts of the review process have examined content validity further from a theoretical point of view.

A1.2.3 Construct validity

Statistical evaluation is a key component of construct validation. Construct validity occurs where an instrument behaves in a way that could be predicted by an underlying theory. Statistical assessment of construct validity is concerned with the structure of an instrument, its correlations with theoretically related and unrelated phenomenon (*eg* harm). This last form of construct validity subsumes concurrent validity, predictive validity and discriminant validity.

A1.2.4 Construct validity and factor analysis

Construct validity as determined by an instrument's structure: An instrument should have a theoretically interpretable factor structure. For instance, a scale that is thought to be uni-dimensional should consist of only one factor is subjected to factor analysis; a scale that is thought to be multi-dimensional should produce as many factors as there are sub-scales from factor analysis, items should load on their hypothesised scales in the expected direction (>.30 or better) and items loading on non-hypothesised scale be <.30.

A1.2.5 Factor analysis

There are two main forms of factor analysis, exploratory factor analysis (EFA, this sub-set also includes the related and commonly used technique of principal components analysis [PCA]) and confirmatory factor analysis (CFA). Each is suitable at different stages of scale development and validation. Whichever technique is used however, it is important that all items in an instrument are analysed together. Some researchers analyse sub-scales separately to demonstrate the uni-dimensionality of each scales. This is not acceptable: in so doing, such researchers are not testing the overall structure and hence the validity of the whole instrument. Further, the approach also prevents detection of non-hypothesised cross-loadings for some items, so making it more likely that 'noisy' or inaccurate items will be retained in the final instrument.

A1.2.6 Exploratory factor analysis

Exploratory factor analysis is most suitable in the early stages of scale development (Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg and Williams, 1997). There are a number of reasons for this, this most important being:

- Unlike CFA, EFA is able to uncover non-hypothesised cross-loadings.
- The provision of eigenvalues in EFA provides direct diagnostic information on the number of factors underlying the data.

Therefore, it is preferable if the initial or early factor analytic studies of an instrument contain EFA of the instrument (well established instruments are better subjected to CFA, see below).

A number of variants of EFA are suitable, but the most common are PCA and principal axis factoring (PAF). Arguably, PCA provides an empirical summary of a given set of data, making it less suitable than 'true' factor analysis, such as PAF, for theoretical problems (Tabachnik and Fiddel, 1989), but in many situations this may be a case of 'splitting hairs'. It is usual to expect factors to

correlate, so oblique rotation is often used (usually OBLIMIN), but in some circumstance orthogonal rotation might be used instead (often VARIMAX). Practical experience indicates that quibbling over rotation choices here may be 'splitting hairs' too in many situations.

There are a number of indicators of the suitability of a data set that are usually reported in papers (see, for example Norusis, 1988, Tabachnik and Fiddel, 1989). These are:

a) Sample size: Ideally the sample size should exceed four times the number items in a scale or 100, whichever is the greater *ie* 100 is about the minimum required sample for a factor analysis. If there are more than 25 items in the scale the number of respondents should be at least four times the number of items. Generally speaking, the greater the sample size the more stable the results. Thus, sample sizes exceeding ten times the number of items in a scale or 200, whichever is the greater, are likely to give more stable solutions. Sample sizes exceeding 20 times the number of items in a scale or 400, whichever is the greater, are likely to give even more stable solutions. Sample sizes exceeding 1,000 are likely to give the most stable solutions.

b) The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is also an index of the size of correlations amongst items. This should be reported and should exceed 0.70. Values in excess of 0.80 or even 0.90 are preferred.

The adequacy of fit of an EFA can be judged in several ways. These are:

a) *Extraction of the correct number of factors*. There are several decision rules here (see for example Glorfeld, 1995 for a review). The minimum criterion is that only factors with eigenvalues > 1 are extracted. Since this tends to extract too many factors in many instances, Cattell's scree plot is often also used as a diagnostic tool. This decision rule can be liberal, so researchers sometimes use more conservative ones, such as eigenvalue >1.5 or >2; parallel analysis (where the scree plot is compared to a scree plot that would be expected by chance); or, where large sample sizes permit, cross-validation of the factor solution in two or more sub-samples. Each of these approaches has merit, but perhaps the most useful decision rule is to extract the number of factors that produce the solution that makes the most theoretical sense, provided other statistical decision rules are taken into account.

b) *Variance accounted for by the factors*. Ordinarily, the factors should account for a large proportion of variance shared by the items (>70 per cent) post-rotation. In practice, this might be difficult to achieve, so factors that account for >50 per cent (or even approaching 50 per cent) of the variance amongst items may be considered acceptable. (Note, if the first factor accounts >40 per cent variance, it is possible that the scale is uni-dimensional,

which for a scale suspected to be multi-dimensional may indicate substantial response bias).

c) *Items load on expected factors, or item loadings make theoretical sense*. Factor loadings >.30 are usually taken as the minimum threshold for a significant factor loading, but loadings of >.40 or >.50 might be acceptable. It is important that the pattern of loadings indicates a theoretically interpretable factor.

d) *Items do not have large cross-loadings on several factors*. Items with cross-loadings on several factors are 'noisy' or inaccurate. Retaining them can compromise the validity of the instrument.

A1.2.7 Confirmatory factor analysis

Confirmatory factor analysis, using software programmes such as LISREL, EQS or AMOS, is usually applied in replication studies of an instrument's validity. As noted above, there are obvious advantages of using EFA first. CFA should only be used in the initial stages of instrument development when a) there is a *very* strong theoretical basis for the instrument and/or b) where there is a strong reason to suspect a complex factor structure involving response bias or complex error terms (Daniels, 2000). CFA can model *a priori* structures, including *a priori* specification of method factors or complex error terms; EFA cannot do this. Therefore, where CFA is used, an a priori structure or structure(s) should be specified, either on the basis of strong theory or prior development of a scale. CFA's power is this *ability to test predicted models*. It is then preferable that CFA be conducted at some point in an instrument's development.

There are no definitive guidelines for the minimum sample size for CFA, although 200 seems widely accepted as the minimum feasible to draw accurate inference (Boomsma, 1982). As with EFA though, a similar set of sample size guidelines can be drawn up: the minimum sample size should exceed four times the number of items in a scale or 200, whichever is the greater. Sample sizes exceeding ten times the number of items in a scale or 200, whichever is the greater, are likely to give more stable solutions. Sample sizes exceeding 20 times the number of items in a scale or 400, whichever is the greater, are likely to give even more stable solutions. Sample sizes exceeding 1,000 are likely to give the most stable solutions.

CFAs should also be conducted on the covariance rather than the correlation matrix, as recommended by Cudeck (1989).

It is possible, especially in the absence of strong theory, that where an initial validation study has a large sample, the sample can be split in two and EFA conducted on one half, and then CFA fitted to the other half, using the pattern of loadings or exact loadings found in EFA. Although this practice is laudable, it is however *desirable* if CFAs are also reported in separate studies. This is

because CFA on the other half of these data would be collected by the same researchers, using the same protocols and procedures as for the data subjected to EFA. Therefore, there are likely to be common errors across both samples, making true cross-validation problematic (Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg and Williams, 1997). Even independent samples gathered by the same group of researchers may share common problems across studies, making independent validation using CFA on new samples also desirable.

Whilst FFA is more sensitive to departures from the assumption of multivariate normality than CFA if maximum likelihood estimation is used (Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg and Williams, 1997), the provision of robust statistics in programmes such as EQS or the use of alternative estimation with less restrictive assumptions such as arbitrary distribution generalised least squares helps circumvent this problem (Dunn, Everitt and Pickles, 1993). Further, deviation from multivariate normality does not appear to affect CFA unduly, except where skew is very large (Harlow, 1985).

There are several ways to check the fit of an instrument to its hypothesised structure. The most important are listed below:

Use of fit indices. Most CFA packages give several fit indices. In many cases, fit indices such as the significance of χ^2, AIC (Akaike's information criterion) and CAIC (Bozdogan's variant of Akaike's information criterion) give values specific to that sample, and should be used only to compare alternative models fitted on the same sample for the instrument. There are several indices whose range is ordinarily approximately between 0 and 1 for all samples, and definitive guidelines can be given for judging the fit of a hypothesised model across samples. These fit indices include (see for example Medsker, Williams and Holahan, 1994):

i) The Goodness of Fit Index, which should exceed 0.90 for adequate fit.

ii) The Adjusted Goodness of Fit Index, which should exceed 0.90 for adequate fit.

iii) The Tucker-Lewis Index (TLI) or Non-Normed Fit Index (NNFI), which should exceed 0.90 for adequate fit.

iv) The Normed Fit Index (NFI), which should exceed 0.90 for adequate fit.

v) The Parsimonious Fit Index (PFI or PNFI), which should exceed 0.90 for adequate fit.

vi) The Comparative Fit Index (CFI), which should exceed 0.95 for adequate fit.

Note that all fit indices should exceed 0.90, but the current consensus is that the CFI should exceed 0.95 (Hu and Bentler,

1998), although previously values of >.90 were considered adequate for the CFI.

Importantly, it is recommended that researchers report several fit indices (Medsker, Williams and Holahan, 1994), otherwise we would be unsure whether researchers are reporting the most favourable fit index only.

Significant loadings — CFA provides significant tests of factor coefficients. All hypothesised loadings then should be statistically significant (usually p<.05 or better) and in the hypothesised direction.

Absence of post hoc modification or *cross-validation* in a separate sample. CFA provides diagnostics for improving model fit (*eg* Lagrange multiplier test). Sometimes, researchers may conclude that the best fitting factor structure for an instrument is a model derived from application of *post hoc* modification tests. This effectively makes CFA more like EFA, and the results may reflect sample bias rather than true misspecification of the original model. Ideally, the model chosen as fitting the data should be hypothesised *a priori* on the basis of strong theory. Where modification tests are used, the results should be cross-validated either on a holding sample from the original sample pool or a separate sample.

A1.2.8 Construct validity as concurrent validity

Concurrent validity is demonstrated if an instrument or sub-scales of an instrument are associated with theoretically related variables. This is often inferred from patterns of significant correlations with variables measuring indices of harm; such as health outcomes, psychological well-being and job satisfaction. More rarely, other indices are taken, such as measures of performance, absenteeism or coping. In addition, instruments that assess hazards should be related to job descriptions or job categories in a way that would be expected. It is desirable that the pattern of relationships is different for each sub-scale in the sample, to be sure that each sub-scale is not measuring the same underlying construct (perhaps a response style or underlying personality trait such as negative affectivity). Ideally, the pattern of correlations and relations should be specified *a priori*, but it is more important that the patterns make theoretical sense.

A1.2.9 Construct validity as predictive validity

The same principles can be applied to predictive validity as for concurrent validity, excepting that with predictive validity sub-scales of an instrument should be associated significantly with measures of harm or other dependent variable taken at some point after the measurement of hazards. Ideally, the pattern of associations should control for levels of harm or other dependent

variable taken before or concurrently with the measures of hazards, so that the measures of psychosocial hazards are predicting future changes in harm. Given the basic purpose of hazard assessment – to measure features of work that may cause harm, predictive validity is extremely important.

A1.2.10 Construct validity as discriminant validity

Discriminant validity may be assessed with measures taken concurrently or subsequent to measures of hazards. Discriminant validity is said to exist where the sub-scales of an instrument *are not* significantly associated with theoretically unrelated variables (*eg* social desirability) (Campbell and Fiske, 1959).

It is currently contentious whether there should be no association between hazard measures and measures of negative or positive affectivity (Spector, Zapf, Chen and Frese, 2000). However, at a minimum, associations between an instrument's sub-scales and harm should remain significant, after controlling for measures of negative affectivity in partial correlation analysis. Not all research reports may report such partial correlations, although they can be worked out by hand from zero-order correlation matrices (see Hays, 1988).

A1.3 Psychometric criteria for single studies

The following criteria were produced to assist reviewers evaluating each paper. These ratings were used along with additional reviewer comments to evaluate each report. Where possible, an overall summative assessment was made for each instrument on the basis of the assessments for each study reviewed (criteria to assist in this are outlined in the next section).

A1.3.1 Reliability

Internal consistency

A. Acceptable: All sub-scales reliable (alphas or split half reliabilities >.70).

B. Approaching acceptability: Most sub-scales reliable (r>.70), some marginal (>.60).

C. Unacceptable: Instrument unreliable (many or all rs <.70).

n/a — single item scale.

n/r — internal consistency not reported.

Test-retest reliability

A. Acceptable: Test-retest reliability good for all sub-scales r>.70 over short and stable period).

B. Approaching acceptability: Most sub-scales reliable (r>.70), some marginal (>.60), over short and stable period.

C. Unacceptable: Instrument unreliable (many or all rs <.70, over short and stable period).

n/a (1) — only one wave of data collected.

n/a (2) — period of months or years elapses between measurements.

n/r — test-retest reliability not reported.

NB — a short and stable period is defined as a few days to a few weeks, where no organisational or job change takes place.

Test-retest sensitivity

A. Good: Test-retest reliabilities <.70 for all scales over long or unstable period.

B. Marginal: Some test-retest reliabilities <.70 over long or unstable period.

C. Poor: All test-retest reliabilities >.70 over long or unstable period.

n/a (1) — only one wave of data collected.

n/a (2) — period between measurement = few days — weeks, where no organisational or job change takes place.

n/r — test-retest reliability not reported.

NB — a long period is defined as several months to several years. An unstable period is defined as occurring where job or organisational change has occurred in the interval between measurements.

Inter-rater reliability

A. Acceptable: All sub-scales reliable (inter-rater rs >.70).

B. Approaching acceptability: Most sub-scales reliable (r>.70), some marginal (>.60).

C. Unacceptable: Instrument unreliable (many or all rs <.70).

n/a — self-report instruments.

n/r — single raters used or inter-rater reliability not reported.

NB: Intra-class correlation coefficients can be artefactually low, and their use should be noted.

Response rates with instrument

A. Very good > 60 per cent.

B. Good > 50 per cent.

C. Fairly good > 40 per cent.

D. OK > 30 per cent.

E. Poor > 20 per cent.

F. Very poor < 20 per cent.

(NB — these figures were developed from reading the relevant literature and experience with psychosocial hazard measures, rather than formal statistical criteria.).

A1.3.2 Validity

Face validity

A. Likely to be good (items developed from interviews etc with target population or pre-tested and revised on the target population).

B. Unknown (no effort to develop or pre-test items on a target population).

Content validity

A. Good: covers full range of phenomena the measure purports to and frequency based response scales with specified time period.

B. Marginal: covers full range of phenomenon the measure purports to or frequency based response scales with specified time period.

C. Poor: neither covers the full range of phenomenon the measure purports to nor uses frequency based response scales with specified time period.

NB: These criteria for content validity are augmented by other theoretical criteria derived by the project team.

A1.3.3 Exploratory factor analysis

n/a — not used.

Use of EFA in early stages of instrument development

A. Yes

B. No.

Sample size

A. Very good: size > 1,000.

B. Good: size > 20 times the number of items in a scale or 400.

C. OK: size > ten times the number of items in a scale or 200.

D. Minimum: size > four times the number items in an instrument or 100, whichever is the greater.

E. Not acceptable: size < four times number of an instrument or < 100.

n/r: not reported.

Keiser Mayer Olkin

A. Marvellous >.90.

B. Meritorious >.80.

C. Middling >.70.

D. Mediocre >.60.

E. Miserable >.50.

F. Unacceptable <.50.

n/r: not reported.

Overall evaluation of suitability of data for EFA.

A. Suitable: Sample size and KMO all fall within A,B,C range.

B. Marginal: Sample size — range A-D, KMO range A-E.

C. Not suitable: sample size = E/nr, E/nr, KMO = F/nr.

Extraction of correct number of factors

A. Highly likely: cross-validation on two samples, use of scree plot, eignenvalue > 1 or other more conservative rules used. factors extracted produce interpretable scales.

B. Likely: Use of scree plot, eignenvalue > 1 or other more conservative rules used, and factors extracted produce interpretable scales.

C. Unlikely: eigenvalue > 1 rule used, and scales not interpretable.

n/d: No decision rules made explicit.

Variance accounted for by all the factors (post-rotation)

A. Very good: > 70 per cent.

B. OK: > 50 per cent.

C. Marginal: > 40 per cent.

D. Poor < 40 per cent. (NB if a single factor extracted and the scale is uni-dimensional, >30 per cent is acceptable).

n/r — variance accounted for not reported.

NB: For principal components analysis, post-rotation variance accounted for = pre-rotation variance accounted for.

Items load on expected factors or pattern of loadings makes theoretical sense

Acceptable: Loadings > |.30|.

Marginal: Most loadings > |.30|, one or two < |.30|.

C. Poor: Loadings < |.30|.

n/r — factor loadings not reported.

Items do not have cross-loadings on several factors

A. Good: no cross-loadings > |.30|.

B. Marginal: One or two cross-loadings > |.30|.

C. Unacceptable: Several cross-loadings > |.30|.

n/r — factor loadings not reported.

Overall evaluation of EFA solution

A. Excellent: factor extraction A; variance accounted for A; loadings A; cross-loadings A.

B. Good: three A and one B from the following: factor extraction A/B; variance accounted for A/B; loadings A/B; cross-loadings A/B.

C. Marginal: factor extraction A/B; variance accounted for A/B; loadings A/B; cross-loadings A/B.

D. Unacceptable: factor extraction B or less; variance accounted for B or less; loadings B or less; cross-loadings B or less.

A1.3.4 Confirmatory factor analysis:

n/a — not used.

Use of CFA

A. Very appropriate: prior EFAs conducted on scales with independent samples or strong *a priori* structure.

B. Appropriate: prior EFAs conducted on scales or strong *a priori* structure.

C. Not appropriate: neither prior EFAs conducted with scales nor strong *a priori* structure.

Sample size

A. Very good: size > 1,000.

B. Good: size > 20 times the number of items in a scale or 400.

C. OK: size > ten times the number of items in a scale or 200.

D. Minimum: size > four times the number items in an instrument or 200, whichever is the greater.

E. Not acceptable: size < four times number of an instrument or < 200.

n/r: not reported.

Covariance matrix

A. Analysis on covariance matrix.

B. Analysis on correlation matrix.

n/r matrix used in analysis not reported.

Evaluation of suitability of data for CFA

A. Suitable: Use of CFA – A/B; sample size A-C, covariance matrix A.

B. Marginal: Use of CFA – A/B; sample size A-D, covariance matrix A/B.

C. Not suitable: Use of CFA – C, sample size D/E, covariance matrix B.

Model fit

A. Good: Several fit indices of kind described, all exceeding minimum threshold.

B. Marginal: Several fit indices of kind described, with all but one exceeding minimum threshold.

C. Uncertain: One fit index of kind described, exceeding minimum threshold.

D. Unacceptable: One fit index reported below minimum threshold, or several indices reported below minimum threshold.

Loadings

A. Acceptable: all significant and in hypothesised direction.

B. Marginal: all but one or two significant and all in hypothesised direction.

C. Unacceptable: several non-significant loadings, or not in hypothesised direction.

Use of modification tests

A. Acceptable: Modification tests not used, or modified scale structure on a separate sample.

B. Unacceptable: Modification tests used, but not attempt at cross-validation in a separate sample.

A1.3.5 Overall evaluation of CFA solution

A. Excellent: Model fit A, loadings A, modification tests, A.

B. Possibly acceptable: Model fit A, loadings A, modification tests, modification tests A/B.

C. Marginal: Model fit A/B, loadings A/B, modification tests A/B.

D. Unacceptable: None of the above.

A1.3.6 Concurrent validity

A. Good: several statistically significant relationships/correlations with theoretically related variables measured at same time, including some measures of harm. Pattern of relationships is different for each sub-scale.

B. Poor: statistically significant relationships/correlations with theoretically related variables measured at same time, including some measures of harm. Pattern of relationships is almost the same for each sub-scale, including correlation coefficients of similar size, or relationships occur with only one or two other variables.

Unacceptable: no statistically significant relationships/ correlations with theoretically related variables measured at same time.

n/a — no correlations with external variables reported.

A1.3.7 Predictive validity

A. Very good: statistically significant relationships/correlations with theoretically related variables measured after administration of hazard measure, including some measures of harm. Pattern of relationships is different for each sub-scale. Relationships remain significant after controlling for initial levels of predicted variable.

B. Good: statistically significant relationships/correlations with theoretically related variables measured after administration of hazard measure, including some measures of harm. Pattern of relationships is different for each sub-scale. Most relationships remain significant after controlling for initial levels of predicted variable.

C. Poor: statistically significant relationships/correlations with theoretically related variables measured after administration of hazard measures, including some measures of harm. Pattern of relationships is almost the same for each sub-scale, including correlation coefficients of similar size, relationships are with only one or two other variables or pattern of significant relationships disappears after controlling for initial levels of predicted variables.

Unacceptable: no statistically significant relationships/ correlations with theoretically related variables measured after administration of hazard measures.

n/a — no predictive correlations reported.

A1.3.8 Discriminant validity

A. Good: No statistically significant relationships with theoretically unrelated variables, or partial correlations between sub-scales of instrument and harm remain significant after controlling for negative affectivity.

B. Questionable: No statistically significant relationships with theoretically unrelated variables, but correlations with measures of negative affectivity not reported.

C. Poor: Statistically significant correlations with theoretically unrelated variables, or partial correlations between sub-scales of instrument and harm are not significant after controlling for negative affectivity.

A1.4 Overall summative assessment across studies

The following criteria were developed to assist the project team in summarising the results of several studies using an instrument and, where possible, to arrive at an overall evaluation of each instrument.

A1.4.1 Overall quality of validation studies

To be counted as a full validation study, at a minimum, the study should include reports of internal consistency, factor analysis and reports of relationships/correlations with theoretically related variables.

1. Very good: Several validation analyses, including several studies conducted by independent research teams.
2. Good: More than one validation study, and at least one validation study conducted by an independent research team.

3. Promising, but additional independent evidence needed. Two or more studies conducted, but all by research teams connected to scale developers.
4. Additional evidence needed. Only one validation study conducted.

A1.4.2 Overall evaluation of internal consistency

5. Acceptable: Internal consistency rated A across all of several studies.
6. Possibly acceptable: Internal consistency mainly As with some Bs across several studies.
7. Problematic: Internal consistency mainly Bs across studies.
8. Not acceptable: Internal consistency Bs or Cs across studies.
9. Additional evidence needed but promising: internal consistency data available from only one study — rated A.
10. Additional evidence needed but problematic: internal consistency data available from only one study — rated B or C.
11. Unknown: no internal consistency data available.

A1.4.3 Overall evaluation of test-retest reliability

12. Acceptable: Test-retest reliability rated A across all of several studies.
13. Possibly acceptable: Test-retest reliability mainly As with some Bs across several studies.
14. Problematic: Test-retest reliability mainly Bs across studies.
15. Not acceptable: Test-retest reliability Bs or Cs across studies.
16. Additional evidence needed but promising: test-retest data available from only one study — rated A.
17. Additional evidence needed but problematic: test-retest data available from only one study — rated B or C.
18. Unknown: no test-retest data available.

A1.4.4 Overall evaluation of test-retest sensitivity

19. Acceptable: Test-retest sensitivity rated A across all of several studies.
20. Possibly acceptable: Test-retest sensitivity mainly As with some Bs across several studies.
21. Problematic: Test-retest sensitivity mainly Bs across studies.
22. Not acceptable: Test-retest sensitivity Bs or Cs across studies.

23. Additional evidence needed but promising: test-retest data available from only one study — rated A.
24. Additional evidence needed but problematic: test-retest data available from only one study — rated B or C.
25. Unknown: no test-retest data available.

A1.4.5 Overall evaluation of inter-rater reliability

26. Acceptable: Inter-rater reliability rated A across all of several studies.
27. Possibly acceptable: Inter-rater reliability mainly As with some Bs across several studies.
28. Problematic: Inter-rater reliability mainly Bs across studies.
29. Not acceptable: Inter-rater reliability Bs or Cs across studies.
30. Additional evidence needed but promising: Inter-rater reliability data available from only one study — rated A.
31. Additional evidence needed but problematic: Inter-rater reliability data available from only one study — rated B or C.
32. Unknown: no inter-rater reliability data available.

A1.4.6 Overall quality of reliability evidence

33. Good. Rated 1 for internal or inter-rater, test-retest reliability and test-retest sensitivity.
34. Promising. Rated 1 for internal or inter-rater reliability, rated 5 for both test-retest reliability and test-retest sensitivity.
35. Possibly acceptable. Rated 1 for internal or inter-rater reliability, 7 for both test-retest reliability and test-retest sensitivity.
36. Problematic. Rated 2 for internal or inter-rater reliability, rated 1 or 5 for test-retest reliability for both test-retest reliability and test-retest sensitivity.
37. Unacceptable. None of the above.

A1.4.7 Overall evaluation of face validity evidence

Response rates with instrument

38. Very good: Response rate rated A across all of several studies.
39. Good: Response rate rated A or B across all of several studies.
40. Fairly good: Response rate rated A, B or C across all of several studies.
41. OK. Response rate rated A-D across all of several studies.

42. Possible problems. Response rate rated A-D across most of several studies, occasional E or F.

43. Poor. Response rate rated mainly E or F across several studies.

Face validity

1. Likely to be good (items developed from interviews *etc.* with target population or pre-tested and revised on target population) — may be conducted over several studies.

2. Unknown (no effort to develop or pretest items on the target population) — if face validity ignored across several studies.

Content validity

1. Good: covers full range of phenomenon the measure purports to and frequency based response scale with specified time period.
2. Marginal: covers full range of phenomenon the measure purports to or frequency based response scale with specified time period.
3. Poor: neither covers the full range of phenomenon the measure purports to nor uses frequency based response scale with specified time period.

NB: These criteria for content validity are augmented by other theoretical criteria derived by the project team.

A1.4.8 Overall quality of validity evidence: factor analysis

44. Good: Scales subject to independent EFA and CFA in different studies, scale structure replicated across studies and EFA and CFA solutions rated A for all studies.

45. Possibly good (I) — needs CFA evidence: Scales subject to EFA in different studies, scale structures replicated across studies, EFA solutions rated A across studies.

46. Possibly good (II) — needs additional studies: Scales subject to EFA and CFA in the study on different samples, scale structure replicated across analyses and EFA solution rated A, CFA solution rated A.

47. Possibly good (III) — needs additional samples: Scales subject to EFA or CFA in one study on one sample, EFA/CFA solution rated A.

48. Problematic (I): Scales subject to independent EFA and CFA in different studies, scale structure replicated across studies and EFA solutions rated A or B across studies, CFA rated A or B across studies.

49. Problematic (II): Scales subject to EFA in different studies, scale structures replicated across studies, EFA solutions rated A or B across studies.
50. Problematic (III): Scales subject to EFA or CFA in one study on one sample, EFA/CFA solution rated B.
51. Not valid: Scales subject to EFA or CFA across several studies, and either structure not replicated, or EFAs or CFAs rated C or worse.
52. Unknown: No EFAs or CFAs reported.

A1.4.9 Overall quality of validation evidence: concurrent, predictive and discriminant

Overall quality of concurrent validity

53. Good: concurrent validity rated A across all of several studies.
54. Possibly acceptable: concurrent validity mainly As with some Bs across several studies.
55. Problematic: concurrent validity mainly Bs across studies.
56. Not acceptable concurrent validity Bs or Cs across studies.
57. Additional evidence needed but promising: Concurrent validity data available from only one study — rated A.
58. Additional evidence needed but problematic: Concurrent validity data available from only one study — rated B or C.
59. Unknown: no concurrent validity data available.

Overall quality of predictive validity

60. Good: predictive validity rated A or B across all of several studies.
61. Possibly acceptable: predictive validity mainly As or Bs with some Cs across several studies.
62. Problematic: predictive validity mainly Cs across studies, with some Bs.
63. Not acceptable predictive validity Cs across studies.
64. Additional evidence needed but promising: Predictive validity data available from only one study — rated A or B.
65. Additional evidence needed but problematic: Predictive validity data available from only one study — rated C.
66. Unknown: no predictive validity data available.

Overall quality of discriminant validity

67. Good: discriminant validity rated A across all of several studies.
68. Possibly acceptable: discriminant validity As or Bs across several studies.
69. Problematic: discriminant validity mainly As or Bs across several studies, with some Cs.
70. Not acceptable: discriminant validity mainly Cs across studies.
71. Additional evidence needed but promising: discriminant validity data available from only one study — rated A.
72. Additional evidence needed but problematic: discriminant validity data available from only one study — rated C.
73. Unknown: no discriminant validity data available.

Overall rating of construct validation evidence

74. Good: Factor analysis, concurrent, predictive and discriminant validity all rated 1.
75. Possibly acceptable: Factor analysis rated 1-4, concurrent, predictive and discriminant validity all rated 1 or 2.
76. Problematic: Factor analysis rated 1-7, concurrent, predictive and discriminant validity all rated 1-3.
77. Very problematic: Factor analysis rated 1-7 or 9, concurrent, predictive and discriminant validity all rated 1-3 or 5-7.
78. Unacceptable: One from the following: Factor analysis rated 8; Concurrent validity rated 4; Predictive validity rated 4; discriminant validity rated 4.

Appendix 2: Single Study Proforma

Paper ID:

SINGLE STUDY PROFORMA

NB Please refer to the notes on statistical review criteria when using this proforma. Where more than one study is reported in an article, follow the order in which they appear in the article, enter the number of the study at the top of each proforma and keep together.

Response rate

N = ☐ Reponse rate = ☐

Reliability

Name of scale	Cronbachs Alpha	Split half Reliability	Test-Retest Reliability	Test-Retest Sensitivity

The reliability of externally rated tools

Inter-rater reliability:

Not rep'd ☐ 9 Not used ☐ 8 N/A ☐ 7 most 'r's>0.7 ☐ 0 most 'r's<0.7 ☐ 1 all 'r's<0.7 ☐ 2

Validity

Concurrent Validity:

Not rep'd ☐ 9 Good ☐ 2 Poor ☐ 1 Unacceptable ☐ 0

Predictive validity:

Not rep'd ☐ 9 V good ☐ 3 Good ☐ 2 Poor ☐ 1 Unacceptable ☐ 0

Divergent Validity:

Not rep'd ☐ 9 Good ☐ 2 Poor ☐ 1 Unacceptable ☐ 0

Construct validity:

Either here or in other work reported in the article, was:

EFA used as the tool was being developed? Yes ☐ 1 No ☐ 0

CFA used as the tool was being developed? Yes ☐ 1 No ☐ 0

NB: if 'Yes' for EFA or CFA refer to statistics expert as the second reviewer.

Appendix 3: Pro Formas for Exploratory and Confirmatory Factor Analyses

Paper ID No. ☐

EXPLORATORY FACTOR ANALYSIS

Please tick/circle as appropriate

1. Use of EFA in early stages of instrument development? yes ☐ 1 no ☐ 2

2. Sample size:

Very good: size > 1000	5
Good: size > 20 * the no. of items in a scale or 400	4
OK: size > 10 * the no. of items in a scale or 200	3
Minimum: size > 4 * the no. items in an scale or 100	2
Not acceptable: size < 4 * no. of an scale or < 100	1
Not reported	9

3. KMO:

Marvellous >.90	6
Meritorious >.80	5
Middling >.70	4
Mediocre >.60	3
Miserable >.50	2
Unacceptable <.50	1
Not reported	9

4. Suitability of data for EFA:

Suitable (*ie* sample size = 3+ and KMO = 4+)	3
Marginal (*ie* sample size = 2+ and KMO = 2+)	2
Not suitable (*ie* sample size and/or KMO = 1 or 9)	1

5. Extraction of correct no. of factors:

No decision rules made explicit.	9
Highly likely: cross-validation on two samples, + scree plot, eigenvalue > 1 or other more conservative rules used. Factors extracted produce interpretable scales.	3
Likely: Scree plot, eigenvalue > 1 or other more conservative rules used, + factors extracted produce interpretable scales.	2
Unlikely: eigenvalue > 1 rule used, and scales not interpretable.	1

6. Variance accounted for by all the factors:

Variance not reported	9
Very good: > 70%	4
OK: > 50%	3
Marginal: > 40%	2
Poor < 40%	1

NB If a single factor is extracted and the scale is uni-dimensional, >30% is acceptable. For principal components analysis, post-rotation variance accounted for = pre-rotation variance accounted for.

7. Items load on expected factors or pattern of loadings makes theoretical sense:

Acceptable: Loadings > |.30|	3
Marginal: Most loadings > |.30|	2
Poor: Most loadings < |.30|	1
Factor loadings not reported	9

NB Be wary of loading cut-off criteria exceeding .50 if loadings < .50 are not reported.

8. Items not cross-loaded:

Good: no cross-loadings > |.30|	3
Marginal: One or two cross-loadings > |.30|	2
Unacceptable: Several cross-loadings > |.30|	1
Factor loadings not reported	9

9. Overall evaluation of EFA solution:

Excellent: (top on all ratings)	4
Good: (at least three top ratings)	3
Marginal:(two or fewer top ratings)	2
Unacceptable: (low ratings on all 4)	1

Paper ID No. [ ]

CONFIRMATORY FACTOR ANALYSIS

10. Use of CFA:

Very appropriate: prior EFAs conducted on scales with independent samples or strong *a priori* structure.	3
Appropriate: prior EFAs conducted or strong *a priori* structure.	2
Not appropriate: neither prior EFAs conducted with scales nor strong *a priori* structure	1

11 Sample size:

Very good: size > 1000	5
Good: size > 20 * the no. of items in a scale or 400	4
OK: size > 10 * the no. of items in a scale or 200	3
Minimum: size > 4 * the no. items in an scale or 100	2
Not acceptable: size < 4 * no. of an scale or < 100	1
Not reported	9

12. Covariance matrix:

Matrix used in analysis not reported	9
Analysis on covariance matrix	2
Analysis on correlation matrix.	1

13. Evaluation of suitability for CFA

Suitable: (Use of CFA – 3/2; sample size 5-3, covariance matrix 2)	3
Marginal: (Use of CFA – 3/2; sample size 5-2, covariance matrix 2/1)	2
Not suitable: (Use of CFA – 1; sample size 2/1; covariance matrix 1/9)	1

14. Model fit:

Good: Several fit indices of kind described, all exceed min threshold.	4
Marginal: Several fit indices, all but one exceed min threshold	3
Uncertain: One fit index which exceeds min threshold.	2
Unacceptable: One or several fit indices all below min threshold	1

15. Loadings:

Acceptable-all significant and in hypothesised direction	3
Marginal: nearly all significant and all in hypothesised direction	2
Unacceptable: several non-sig loadings, or not in hyp'd direction.	1

16. Use of modification tests.

Acceptable: Modification tests not used, or modified scale structure on a separate sample.	2
Unacceptable: Modification tests used, but no attempt at cross-validation in a separate sample.	1

17. Overall evaluation of CFA solution.

Excellent: Model fit 4, loadings 3, modification tests, 2.	4
Possibly acceptable: Model fit 4, loadings 3, modification tests 1 or 2	3
Marginal: Model fit 4/3, loadings 3/2, modification tests 1 or 2	2
Unacceptable: None of the above	1

Bibliography

Amick B C, Celetano D D (1991), 'Structural determinants of the psychosocial work environment: introducing technology in the work stress framework', *Ergonomics*, Vol. 34, No 5, pp 625-646

* Anderson W J R, Cooper C L, Willmott M (1996), 'Sources of Stress in the National Health Service: A Comparison of Seven Occupational Groups', *Work and Stress*, Vol. 10, No. 1, pp 88-95

* Andries F, Kompier M A, Smulders P G (1996), 'Do you think that you health or safety are at risk because of your work? A Large European Study on Psychological and Physical Work Demands', *Work and Stress*, Vol. 10, No. 2, pp 104-118

Anon, (1999), 'Stress at work: a survey of 126 employers', *IRS Employee Health Bulletin*, (11, Oct)

* Bacharach S, Bamberger P (1993), 'Causal Models of Role Stressor Antecedents and Consequences: The Importance of Occupational Differences', *Journal of Vocational Behaviour*, Vol. 41, pp 13-34

Baker E, Israel B, Schurman S (1996), 'Role of Control and Support in Occupational Stress: An Integrated Model', *Social Science and Medicine*, Vol. 43, No. 7, pp 1145-1156

* Barnett R C, Brennan R T (1995), 'The Relationship Between Job Experiences and Psychological Distress: A Structural Equation Approach', *Journal of Organizational Behavior*, Vol. 16, p 259-276

* Beehr T A, Walsh J T, Taber T D (1976), 'Relationship of Stress to Individually and Organizationally Valued States: Higher Order Needs as a Moderator', *Journal of Applied Psychology*, Vol. 61, No. 1, pp 41-47

Beehr T A, Jex S M, Stacy B A, Marshall A M (2000), 'Work stressors and coworker support as predictors of individual strain and job performance', *Journal of Organizational Behavior*, Vol. 21, pp 391-405

Bedeian A G, Mossholder K W, Kemery E R, Armenakis A A (1992), Replication Requisites: A Second Look at Klenke-hamel and Mathieu (1990), *Human Relations*, Vol. 45, No. 10

Bhalla S, Jones B, Flynn D M (1991), 'Role Stress Among Canadian White-Collar Workers', *Work & Stress,* Vol. 5, No. 4, pp 289-299

* Biggam R H, Power K G, MacDonald R R (1997), 'Coping with the Occupational Stressors of Police Work: A Study of Scottish Officers', *Stress Medicine*, Vol. 13, pp 109-115

* Bogg J, Cooper C (1995), 'Job Satisfaction, Mental Health, and Occupational Stress Among Senior Civil Servants', *Human Relations*, Vol. 48, No. 3

Boomsma A (1982), 'The robustness of LISREL against small sample sizes in factor analysis models'. In K G Joreskög and H Wold (eds.), *Systems Under Indirect Observation: Causality, Structure, Prediction*, Part 1. Amsterdam: North-Holland Publishing

* Borg V, Kristensen T S, Burr H (2000), 'Work Environment and Changes in Self-Rated Health: A Five Year Follow-up Study', *Stress Medicine*, Vol. 16, pp 37-47

Bosma H, Marmot M G, Hemingway H, Nicholson A C, Brummer E, Stansfeld S A (1997), 'Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study', *British Medical Journal*, Vol. 314:558

* Bosma H, Peter R, Siegrist J, Marmot M (1998), 'Two Alternative Job Stress Models and the Risk of Coronary Heart Disease', *American Journal of Public Health*, Vol. 88, No. 1

* Bosma H, Stansfeld S A, Marmot M G (1998), 'Job Control, Personal Characteristics, and Heart Disease', *Journal of Occupational Health Psychology*, Vol. 3, No. 4, pp 402-409

* Bradley J, Eachus P (1995), 'Occupational Stress Within a UK Higher Education Institution', *International Journal of Stress Management*, Vol. 2, No. 3, pp 145-158

* Breaugh J A, 'The Measurement of Work Autonomy' (1985), *Human Relations*, Vol. 38, No. 6, pp 551-570

* Breaugh J A (1989), 'The Work Autonomy Scales: Additional Validity Evidence', *Human Relations*, Vol. 42, No. 11, pp 1033-1056

* Breaugh J A, Becker A S (1987), 'Further Examination of the Work Autonomy Scales: Three Studies', *Human Relations*, Vol. 40, No. 6, pp 381-400

* Breaugh J A, Colihan J P (1994), 'Measuring Facets of Job Ambiguity: Construct Validity Evidence', *Journal of Applied Psychology*, Vol. 79, No. 2, pp 191-202

Briner R and Rick J (1999), 'Trauma Management vs. Stress Debriefing: What Should Responsible Organisations Do? Paper presented at the British Psychological Society Conference, January, 2000

* Brisson C, Laflamme N, Moisan J, Milot A, Masse B, Vezina M (1999), 'Effect of Family Responsibilities and Job Strain on Ambulatory Blood Pressure Among Whilte-Collar Women', *Psychosomatic Medicine*, 61, pp 197-204

* Broers P, Evers A, Cooper C L (1995), 'Differences in Occupational Stress in Three European Countries', *International Journal of Stress Management*, Vol. 2, No. 4, pp 171-180

Burke R J (1998), 'Job Insecurity in Recent Business School Graduates: Antecedents and Consequences', *International Journal of Stress Management*, Vol. 5, No. 2

Burke R J (1998), 'Work Stressors Among Recent Business School Graduates'' *Stress Medicine*, Vol. 14, pp 83-89

Campbell D T, Fiske D W (1959), 'Convergent and discriminant validation by the multitrait-multimethod matrix', *Psychological Bulletin,* 56, 81-107, 1959

Caplan R D, Cobb S, French J R P Jnr., Harrison R V and Pinneau S R (1975), *Job demands and worker health: main effects and occupational differences*, Institute of Social Research, University of Michigan

Carayon P (1993), 'A Longitudinal Test of Karasek's Job Strain Model among Office Workers', *Work and Stress,* Vol. 7, No. 4, pp 299-314

Carayon P (1993), 'Job Design and Job Stress in Office Workers', *Ergonomics,* Vol. 36, No. 5, pp 463-477

* Carayon P, Zijlstra F (1999), 'Relationships Between Job Control, Work Pressure and Strain: Studies in the USA and in the Netherlands', *Work and Stress*, Vol. 13(1), pp 32-48

* Cartwright S, Cooper C L, Barron A (1996), 'The Company Car Driver; Occupational Stress as a Predictor of Motor Vehicle Accident Involvement', *Human Relations*, Vol. 49, No. 2

* Champoux J E (1991), 'A Multivariate Test of the Job Characteristics Theory of Work Motivation', *Journal of Organizational Behavior*, Vol. 12, pp 431-446

* Cheng Y, Kawachi I, Coakley E H, Schwarz J, Colditz G (2000), 'Association Between Pyschosocial Work Characteristics and Health Functioning in American Women: Prospective Study', *British Medical Journal*, Vol. 320, pp 1432-6

Cohen S, Kamarck T, Mermelstein R (1983), 'A Global Measure of Perceived Stress', *Journal of Health and Social Behavior*, 24,4, pp 385-396

Cook T D, Campbell D T (1976), 'The design and conduct of quasi-experiments and true experiments in field settings', In M D Dunnette (ed.) *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally

* Cooke R A, Szumal J L (1993), 'Measuring Normative Beliefs and Shared Behavioural Expections in Organizations: The Reliability and Validity of the Organizational Culture Inventory', *Psychological Reports*, Vol. 72, pp 1299-1330

* Cooper C L, Bramwell R S (1992), 'A Comparative Analysis of Occupational Stress in Managerial and Shopfloor Workers in the Brewing Industry: Mental Health, Job Satisfaction and Sickness', *Work and Stress*, Vol. 6, No. 2, pp 127-138

* Cooper C L, Bramwell S S (1992), 'Predictive Validity of the Strain Components of the Occupational Stress Indicator', *Stress Medicine*, Vol. 8, pp 57-60

* Cooper C L, Bramwell S S (1998), The Self-Reported Well-Being of Employees Facing Organizational Change: Effects of an Intervention', *Occupational Medicine*, Vol. 48, No. 6, pp 361-368

* Cooper C L, Cartwright S (1997), 'An Intervention Strategy for Workplace Stress', *Journal of Psychosomatic Research*, Vol. 43, No. 1, pp 7-16

* Cooper C L, Clarke D, Rowbottom A M (1999), 'Occupational Stress, Job Satisfaction and Well-Being in Anaesthetists', *Stress Medicine*, Vol. 15, pp 115-126

* Cooper C L, Kirkcaldy B D, Brown J (1994), 'A Model of Job Stress and Physical Health: The Role of Individual Differences', *Personality and Individual Differences*, Vol. 16, No. 4, pp 653-655

* Cooper C L, Williams J (1991), 'A Validation Study of the OSI on a Blue-Collar Sample', *Stress Medicine*, Vol. 7, pp 109-112

Cooper C L, Sloan S J, Williams S (1988), *Occupational Stress Indicator Management Guide,* Oxford, England, NFER-Nelson

Cooper C L and Marshall J (1976), 'Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health', *Journal of Occupational Psychology*, Vol. 49, pp 11-28

* Cordery J L, Sevastos P P (1993), 'Responses to the Original and Revised Job Diagnostic Survey: Is Education a Factor in Responses to Negatively Worded Items?', *Journal of Applied Psychology*, Vol. 78, No. 1, pp 141-143

Cox T (1993), *Stress research and stress management: Putting theory to work,* HMSO, London

Cox T (2000), *Organisational Interventions for Work Stress: A Risk Management Approach*, HSE Contract Research Report 00/286

Cox T and Griffiths A (1996) in M J Schabracq *et al.* (eds) *Handbook of Work and Health Psychology*, John Wiley & Son

Cox T, Griffiths A, Barlowe C, Randall R and Rial-Gonzalez E (2000), *Organisational interventions for work stress: A risk management approach*, HSE Books

Cudeck R (1989), 'Analysis of correlation matrices using covariance structure models', *Psychological Bulletin*, 105, 317-327

* Cunha R C, Cooper C L, Moura M I, Reis M E, Fernandes P (1992), 'Portuguese Version of the OSI: A Study of Reliability and Validity', *Stress Medicine*, Vol. 8, pp 247-251

Daniels K (2000), 'Measures of five aspects of affective well-being at work', *Human Relations,* 53, 275-294

Daniels K, de Chernatony L, Johnson G (1995), 'Validating a method for mapping managers' mental models of competition', *Human Relations*, 48, 975-991

* Daniels K and Guppy A (1994), 'Occupational Stress, Social Support, Job Control and Psychological Well-Being', *Human Relations*, Vol. 47, No. 12

* Davis A J (1996), 'A Re-analysis of the Occupational Stress Indicator', *Work and Stress*, Vol. 10, No. 2, pp 174-182

* Day D V, Bedeian A G, Conte J M (1998), 'Personality as Predictor of Work-Related Outcomes: Test of a Mediated Latent Structural Model', *Journal of Applied Social Psychology*, Vol. 28, No. 22, pp 2068-2088

* De Croon E M, Van Der Beek A J, Blonk R W, Frings-Dresen M H (2000), 'Job Stress and Psychosomatic Health Complaints

Among Dutch Truck Drivers: A Re-Evaluation of Karasek's Interactive Job Demand-Control Model', *Stress Medicine*, Vol. 16, pp 101-107

* de Jonge J, Bosma H, Peter R, Siegrist J (2000), 'Job Strain, Effort-Reward Imbalance and Employee Well-Being: a Large-Scale Cross-Sectional Study', *Social Science and Medicine*, Vol. 50, pp 1317-1327

* de Jonge J, Janssen P P, Van Breukelen G J (1996), 'Testing the Demand-Control-Support Model among Health-Care Professionals: A Structural Equation Model', *Work and Stress*, Vol. 10, No. 3, pp 209-224

* de Jonge J, Kompier M A (1997), 'A Critical Examination of the Demand-Control-Support Model from a Work Psychological Perspective', *International Journal of Stress Management*, Vol. 4, No. 4

* de Jonge J, Mulder M J, Nijhuis F J (1999), 'The Incorporation of Different Demand Concepts in the Job Demand-Control Model: Effects on Health Care Professional', *Social Science and Medicine*, Vol. 48, pp 1149-1160

* de Jonge J, Reuvers M M, Houtman I L, Kompier M A (2000), 'Linear and Non-linear Relations Between Psychosocial Job Characteristics, Subjective Outcomes, and Sickness Absence: Baseline Results from SMASH', *Occupational Health Psychology*, Vol. 5(2), pp 256-268

* de Jonge K, Van Breukelen G J P, Landeweerd J A, Nijhuis F J N (1999), 'Comparing Group and Individual Level Assessments of Job Characteristics in Testing the Job Demand-Control Model: A Multilevel Approach', *Human Relations*, Vol. 52, No. 1

Dekker I, Barling J (1995), 'Workforce Size and Work-Related Role Stress', *Work and Stress,* Vol. 9, No. 1, pp 45-54

* de Rijk A E, Le Blanc P M, Schaufeli W B (1998), 'Active Coping and Need for Control as Moderators of the Job Demand-Control Model: Effects on Burnout', *Journal of Occupational and Organizational Psychology*, Vol. 71, pp 1-18

Derogatis L R (1978), *Psychological Medicine,* Vol. 8, pp 605

Derogatis L R (1997), 'The Derogatis Stress Profile. A theory driven approach to stress measurement' in Zalaquett C P and Wood J R (eds), *Evaluating Stress, A book of resources*, pp 113-140

* Dewe P (1991), 'Measuring Work Stressors: The Role of Frequency, Duration, and Demand', *Work and Stress*, Vol. 5, No. 2, pp 77-91

Dewe P J (1992), 'Applying the Concept of Appraisal to Work Stressors: Some Exploratory Analysis', *Human Relations,* Vol. 45, No. 2

Dewe P, Brook R (2000), 'Sequential Tree Analysis of Work Stressors: Exploring Score Profiles in the Context of the Stressor-Stress Relationship', *International Journal of Stress Management*, Vol. 7, No. 1

* Dodd N G, Ganster D C (1996), 'The Interactive Effects of Variety, Autonomy and Feedback on Attitudes and Performance', *Journal of Organizational Behavior*, Vol. 17, pp 329-347

Dunn G, Everitt B, Pickles A (1993), 'Modelling co-variances and latent variables using EQS', *London: Chapman Hall*

* Evans B K, Fischer D G (1992), 'A Hierarchical Model of Participatory Decision-Making, Job Autonomy, and Perceived Control', *Human Relations*, Vol. 45, No. 11, pp 1169-1189

* Evers A, Frese M, Cooper C L (2000), 'Revisions and Further Developments of the Occupational Stress Indicator: LISREL results from four Dutch studies', *Journal of Occupational and Organizational Psychology*, Vol. 73, pp 221-240

* Fay D, Sonnentag S, Frese M (1998) (ed. Cooper C L), 'Stressors, Innovation, and Personal Initiative: Are Stressors Always Detrimental?, Theories of Organisational Stress', *Oxford University Press*, Chapter 8, pp 170-189

Ferrie J E, Shipley M J, Marmot M G, Stansfeld S, Davey Smith G (1995), 'Health Effects of Anticipation of Job Change and non-employment: longitudinal data from the Whitehall II study', *British Medical Journal*, Vol. 311, pp 1264-1269

* Fletcher B C, Jones F (1993), 'A Refutation of Karasek's Demand — Discretion Model of Occupational Stress with a Range of Dependent Measures', *Journal of Organizational Behavior*, Vol. 14, pp 319-330

* Fogarty G J, Machin M A, Albion M J, Sutherland L F, Lalor G I, Revitt S (1999), 'Predicting Occupational Strain and Job Satisfaction: The Role of Stress, Coping, Personality, and Affectivity Variables', *Journal of Vocational Behaviour*, Vol. 54, pp 429-452

* Fotinatos-Ventouratos R, Cooper C L (1998), 'Social Class Differences and Occupational Stress', *International Journal of Stress Management*, Vol. 5, No. 4

Fox M L, Dwyer D J, Ganster D C (1993), 'Effects of Stressful Job Demands and Control on Physiological and Attitudinal Outcomes in a Hospital Setting', *Academy of Management Journal*, Vol. 36, No. 2, pp 289-318

French J R P, Kahn R L, Mann F C (1962), 'Work, Health and Satisfaction', *The Journal of Social Issues*, Vol. XVIII, No. 3

French J R P and Kahn R L (1962), 'A programmatic approach to studying the industrial environment and mental health', *Journal of Social Issues,* Vol. 18 pp 1-47

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 C L Cooper and R Payne (eds), *Causes, Coping and Consequences of Stress at Work,* Chichester: Wiley

Frese M (1985), 'Stress at work and psychosomatic complaints — a causal interpretation', *Journal of Applied Psychology,* 70, 2, pp 314-328

* Fried Y (1991), 'Meta-Analytic Comparison of the Job Diagnostic Survey and Job Characteristics Investory as Correlates of Work Satisfaction and Performance', *Journal of Applied Psychology*, Vol. 76, No. 5, pp 690-697

Frone M R, Russell M, Cooper M L (1997), 'Relation of work-family conflict to health outcomes: A four-year longitudinal study of employed parents', *Journal of Occupational and Organizational Psychology,* Vol. 70, pp 325-335

Garst, H, Frese M, Molenaar P C M (2000), 'The Temporal Factor of Change in Stressor-Strain Relationships: A Growth Curve Model on a Longitudinal Study in East Germany', *Journal of Applied Psychology,* Vol. 85, No. 3, pp 417-438

Glorfeld L W (1995), 'An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain', *Educational and Psychological Measurement*, 55, 377-393

* Gonzalez-Roma V, Lloret S (1998), 'Construct Validity of Rizzo *et al.'s* (1970) Role Conflict and Ambiguity Scales: A Multisample Study', *Applied Pyschology: An International Review*, Vol. 47 (4), pp 535-545

* Grady G F, Judd B B, Javian S (1990), 'The Dimensionality of Work Autonomy Revisited', *Human Relations*, Vol. 43, No. 12, pp 1219-1228

* Greelund K J, Liu K, Knox S, McGreath H, Dyer A R, Gardin J (1995), 'Psychosocial Work Characteristics and Cardiovascular Disease Risk Factors in Young Adults: The Cardia Study', *Social Science and Medicine*, Vol. 41, No. 5, pp 717-723

Hackman J R, Oldham G R (1975), 'Development of the Job Diagnostic Survey', *Journal of Applied Psychology*, Vol. 60, No. 2, pp 159-170

* Hackman J R, Oldham G R (1975), 'General Job Satisfaction', 'The Experience of Work: A Compendium and Review of 249 Measures and their Use', (eds J D Cook, S J Hepworth, R D Wall, P B Warr), *Occupational and Organizational Psychology*, 2.6 pp 27-28

* Hackman J R, Oldham G R (1975), *Job Diagnostic Survey: Short Form*, Yale University/University of Illinois

* Hackman J R and Oldham G R (1975), 'Development of the Job Diagnostic Survey', *Journal of Applied Psychology,* Vol. 60, 2, pp 159-170

* Hall J K, Spector P E (1991), 'Relationships of Work Stress Measures for Employees with the Same Job', *Work and Stress*, Vol. 5, No. 1 pp 29-35

Harlow L L (1985), 'Behavior of some elliptical theory estimators with nonnormal data in covariance structures framework: a Monte Carlo study' *Unpublished PhD thesis*, University of California, Los Angeles

* Harris M M, Bladen A (1994), 'Wording Effects in the Measurement of Role Conflict and Role Ambiguity: A Multitrait-Multimethod Analysis', *Journal of Management*, Vol. 20, No. 4, pp 887-901

* Harrison D A, McLaughlin M E (1996), 'Structural Properties and Psychometric Qualities of Organizational Self-Reports: Field Tests of Connections Predicted by Cognitive Theory', *Journal of Management*, Vol. 22, No. 2 pp 313-338

* Hawkins N G, Davies R, Holmes T H (1957), 'Evidence of Psychosocial Factors in the Development of Pulmonary Tuberculosis', *American Review of Tubercular Pulmonary Disease*, Vol. 75, pp 768-780

* Haynes C E, Wall R D, Bolden R I, Stride C, Rick J (1999) 'Measures of Perceived Work Characteristics for Health

Services Research: Test of a Measurement Model and Normative Data', *British Journal of Health Psychology*, Vol. 4, pp 257-275

Hays W L (1988), *Statistics*, 4th ed., New York: Holt, Rinehart and Winston

Health and Safety Commission (1997), *Violence and Agression to Staff in Health Services: Guidance on Assessment and Management* (2nd edition), Health Services Advisory Committee

Health and Safety Executive (1998), 'Five Steps to Risk Assessment', HMSO, London

Health and Safety Executive (1998), 'The Control of Substances Hazardous to Health Regulations', HMSO, London

Health and Safety Executive (1990), 'The Control of Substances Hazardous to Health (Amendment) Regulations', HMSO, London

Heinisch D A, Jex S M (1997), 'Negative Affectivity and Gender as Moderators of the Relationship between Work-Related Stressors and Depressed Mood at Work', *Work & Stress,* Vol. 11, No. 1, pp 46-57

* Holmes T H, Rahe R H (1967), 'The Social Readjustment Rating Scale', *Journal of Psychosomatic Research*, Vol. 11, pp 213-218

* House R J, Rizzo J R (1972a), 'Towards the Measurement of Organizational Practices: Scale Development and Validation', *Journal of Applied Psychology*, Vol. 56, No. 5, pp 388-396

House S J, McMichael A J, Wells J A, Kaplan B H, Landerman L R (1979), 'Occupational Stress and Health among Factory Workers', *Journal of Health and Social Behavior*, Vol. 20 pp 139-160

House R J and Rizzo R (1972b), 'Role conflict and ambiguity as critical variables in a model of organizational behavior', *Organizational Behavior and Human Decision Processes,* Vol. 7, 3, pp 467-505

Hu L T, Bentler P M (1998), 'Fit indices in covariance structure modelling: sensitivity to underparameterized model misspecification', *Psychological Methods*, 3, 424-453

Hurley A E, Scandura T A, Schriesheim C A, Brannick M T, Seers A, Vandenberg R J, Williams L J (1997), 'Exploratory and confirmatory factor analysis: guidelines, issues and alternatives', *Journal of Organizational Behavior*, 18, 667-683

* Hurrell Jr J J, McLaney M A (1988), 'Exposure to Job Stress — A New Psychometric Instrument', *Scandinavian Journal of Work Environment and Health*, Vol. 14. pp 27-28

* Ingeldew D K, Hardy L, Cooper C L (1992), 'On the Reliability and Validity of the Locus of Control Scale of the Occupational Stress Indicator, *Personality and Individual Difference*, Vol. 13, No. 11 pp 1183-1191

IRS Employee Health Bulletin (1999), 'Stress at Work: a Survey of 126 Employers', Vol.11, October

Ivancevich J M, Matteson M T, Freedman S M and Phillips J S (1990), Worksite stress management interventions. American Psychologist, 45, 252-261

Ivancevitch J M, Matteson M T (1980), *Stress at Work,* Glenview, IL, Scott Fresman

* Iwata N, Kawakami N, Haratani T, Murata K, Araki S (1999), 'Job Stressor-Mental Health Associations in a Sample of Japanese Working Adults: Artifacts of Positive and Negative Questions?', *Industrial Health*, Vol. 37(2), pp 263-270

* Jackson P R, Wall T D, Martin R, Davids K (1993), 'New Measures of Job Control, Cognitive Demand and Production Responsibility', *Journal of Applied Psychology*, Vol. 78, No. 5, pp 753-762

James L R, Demaree R G, Wolf G (1984), 'Estimating within-group interrater reliability with and without response bias', *Journal of Applied Psychology*, 69, 85-98

James L R, Demaree R G, Wolf G (1993), 'r_{wg}: an assessment of within-group rater agreement', *Journal of Applied Psychology*, 78, 306-309

Jex S M, Gudanowski D M (1992), 'Efficacy Beliefs and Work Stress: An Exploratory Study', *Journal of Organizational Behavior,* Vol. 13, pp 509-517

* Jex S M, Spector P E (1996), 'The Impact of Negative Affectivity on Stressor-Strain Relations: a Replication and Extension', *Work and Stress*, Vol. 10, No. 1, pp 36-45

* Jex S M, Elacqua T C (1999), 'Self-Esteem as a Moderator: A Comparison of Global and Organization-Based Measures', *Journal of Occupational and Organizational Psychology*, Vol. 72, pp 71-81

* Jones F, Bright J E, Searle B, Cooper L (1998), 'Modelling Occupational Stress and Health: The Impact of the

Demand-Control Model on Academic Research and on Workplace Practice', *Stress Medicine*, Vol. 14, pp 231-236

* Kahn H, Cooper C L (1996), 'How Foreign Exchange Dealers in the City of London Cope with Occupational Stress', *International Journal of Stress Management*, Vol. 3, No. 3, pp 137-145

* Kanner A D, Coyne J C, Schaefer C, Lazarus R S (1981), 'Comparison of Two Modes of Stress Measurement: Daily Hassles and Uplifts Versus Major Life Events', *Journal of Behavioral Medicine*, Vol. 4, No. 1

* Karasek Jr R A (1979), 'Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign', *Administrative Science Quarterly*, Vol. 24

Karasek R, Theovell J (1990), 'Healthy Work, Stress, Productivity, and the Reconstruction of Working Life', *Basic Books Inc.*, New York

* Karasek R (1990), 'Lower Health Risk with Increased Job Control Among White Collar Workers', *Journal of Organizational Behavior*, Vol. 11, p 171-185

* Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B (1998), 'The Job Content Questionnaire (JCQ): An Instrument for Internationally Comparative Assessments of Psychosocial Job Characteristics', *Journal of Occupational Health Psychology*, Vol. 3 No. 4, pp 322-355

Karasek R A (1985), *Job content questionnaire and users' guide*, Los Angeles: University of Southern California Department of Industrial and Systems Engineering

* Kayn H, Cooper C L (1991), 'A Note of the Validity of the Mental Health and Coping Scales of the Occupational Stress Indicator', *Stress Medecine*, Vol. 7, pp 185-187

Keashly L, Hunter S, Harvey S (1997), 'Abusive Interaction and Role State Stressors: Relative Impact on Student Residence Assistant Stress and Work Attitudes', *Work and Stress,* Vol. 11, No. 2, pp 175-185

* Kelloway E K, Barling J (1990), 'Item Content Versus Item Wording: Disentangling Role Conflict and Role Ambiguity', *Journal of Applied Psychology*, Vol. 75, No. 6, pp 738-742

* Kelloway E K, Barling J (1991), 'Job Characteristics, Role Stress and Mental Health', *Journal of Occupational Psychology*, Vol. 64, pp 291-304

Kelly C, Sprigg C and Sreenivasan B, (1998), 'SME Managers' Perceptions of Work Related Stress', *Health and Safety Laboratory Report EWP/15/98*

* Kirkcaldy B D, Cooper C L, Brown J M, Athanasou J A (1994), 'Job Stress and Health Profiles of Smokers, Ex-Smokers and Non-Smokers', *Stress Medicine*, Vol. 10, pp 159-166

* Kirkcaldy B D, Cooper C L, Furnham A F (1999), 'The Relationship between Type A, Internality-Externality, Emotional Distress and Perceived Health', *Personality and Individual Differences*

* Kirkcaldy B, Furnham A, Cooper C L (1994), 'Police Personality, Job Satisfaction and Health', *Studia Psychologica*, Vol. 36, No. 1, pp 55-63

Kirkcaldy B D, Martin T (2000), 'Job Stress and Satisfaction among Nurses: Individual Differences', *Stress Medicine,* Vol. 16, pp 77-89

Kivimaki M, Vahtera J, Pentti J, Ferrie J E (2000), 'Factors Underlying the Effect of Organisational Downsizing on Health of Employees: Longitudinal Cohort Study', *British Medical Journal,* Vol. 320, pp 971-975

Koniarek J, Dudek B (1996), 'Social Support as a Buffer in the Stress-Burnout Relationship', *International Journal of Stress Management,* Vol. 3, No. 2

* Kyrouz E M, Humphreys K (1997), 'Do Health Care Workplaces Affect Treatment Environments?', *Journal of Community and Applied Social Psychology*, Vol. 7(2), pp 105-118

* Lancaster R J, Pilkington A, Graveling R (1999), 'Evaluation of the Organisational Stress Health Audit', *HSE Books*

* Landsbergis P, Theorell T (2000), 'Measurement of Psychosocial Workplace Exposure Variables: Self-Report Questionnaires', *Occupational Medicine*, *State of the Art Reviews*, Vol. 15(1), pp 163-188

* Langan-Fox J (1995), 'Occupational Stress in Female and Male Graduate Managers – A Comparative Study', *Stress Medicine*, Vol. 11, pp 157-175

* Langan-Fox J (1996), 'Validity and Reliability of Measures of Occupational and Role Stress Using Samples of Australian Managers and Professionals', *Stress Medicine*, Vol. 12, pp 211-225

* Langan-Fox J, Poole M E (1995), 'Occupational Stress in Australian Business and Professional Women', *Stress Medicine*, Vol. 11, pp 113-122

* Leong C S, Furnham A, Cooper C L (1996), 'The Moderating Effect of Organizational Commitment on the Occupational Stress Outcome Relationship', *Human Relations*, Vol. 49, No. 10

* Lim T K (1995), 'Stress Demands on School Administrators in Singapore', *Work and Stress*, Vol. 9, No. 4 , pp 491-501

* Lu L, Cooper C L, Chen Y C, Ho Hsu C, Hua Li C, Luan Wu H, Bin Shih J (1995), 'Chinese Version of the OSI: A Study of Reliability and Factor Structures', *Stress Medicine*, Vol. 11, pp 149-155

* Lu L, Cooper C L, Chen Y C, Ho Hsu C, Luan Wu H, Bin Shih J, Hua Li C (1997), 'Chinese Version of OSI: A Validation Study', *Work and Stress*, Vol. 11, No. 1, pp 79-86

* Lu L, Tseng H, Cooper C L (1999), 'Managerial Stress, Job Satisfaction and Health in Taiwan', *Stress Medicine*, Vol. 15, pp 53-64

* Lyne K D, Barrett P T, Williams C, Coaley K (2000), 'A Psychometric Evaluation of the Occupational Stress Indicator', *Journal of Occupational and Organizational Psychology*, Vol. 73, pp 195-220

* Margolis B, Kroes W H, Quinn R P (1974), 'Job Stress: An Unlisted Occupational Hazard', *Journal of Occupational Medicine*, Vol. 16, No. 10, pp 659-661

* Mathieu J E (1992), 'Comments Regarding Substantive Findings and the Bedeian and Armenakis Model, and Issues Related to Structural Equation Modeling', *Human Relations*, Vol. 45, No. 10

McClean A A (1979), 'A Method of Self-Assessment', *Work Stress, Addison-Wesley*

* McElfatrick S, Carson J, Annett J, Cooper C, Holloway F, Kuipers E (2000), Assessing Coping Skills in Mental Health Nurses: Is an Occupation Specific Measure Better than a Generic Coping Skills Scale?, *Personality and Individual Differences*, Vol. 28, pp 965-976

Medsker G J, Williams L J, Holahan P J (1994), 'A review of current practices for evaluating causal models in organizational behavior and human resources management research', *Journal of Management*, 20, 439-464

Melamed S, Kushnir T, Meir E I (1991), 'Attenuating the Impact of Job Demands: Additive and Interactive Effects of Perceived Control and Social Support', *Journal of Vocational Behavior,* Vol. 39, pp 40-53

Miller L H & Smith A D (1987), *Stress Audits,* Brookline, MA: Biobehavioral Associates

* Mino Y, Shigemi J, Tsuda T, Yasuda N, Bebbington P (1999), 'Perceived Job Stress and Mental Health in Precision Machine Workers of Japan: a 2 Year Cohort Study', *Occupational and Environmental Medicine*, Vol. 56, pp 41-45

Moos R H (1981), *The Work Environment Scale Manual,* Palo Alto, CA, Consulting Psychologists' Press

Moos R (1994), *The Work Environment Scale Manual*, (third ed.) Palo Alto, CA, Consulting Psychologists' Press

* Morrow G R, Chiarello R J, Derog L R (1978), 'A New Scale for Assessing Patients' Pyschosocial Adjustment to Medical Illness', *Pyschological Medicine*, 8, pp 605-610

Moyle P (1998), 'Longitudinal Influences of Managerial Support on Employee Well-Being', *Work and Stress*, Vol. 12, No. 1 pp 29-49

* Mullarkey S, Jackson P R, Wall T D, Wilson J R, Grey-Taylor S M (1997), 'The Impact of Technology Characteristics and Job Control on Worker Mental Health', *Journal of Organizational Behavior*, Vol. 18, pp 471-489

Muntaner C, O'Campo P J (1993), 'A Critical Appraisal of the Demand/Control Model of the Psychosocial Work Environment: Epistemological, Social, Behavioral and Class Considerations', *Soc. Sci. Med.*, Vol. 36, No. 11, pp. 1509-1517

* Muntaner C, Eaton W W, Garrison R (1993), 'Dimensions of the Psychosocial Work Environment in a Sample of the US Metropolitan Population', *Work and Stress*, Vol. 7(4), pp 351-363

* Munz D C, Huelsman T J, Konold T R, McKinney J J (1996), 'Are There Methodological and Substantive Roles for Affectivity in Job Diagnostic Survey Relationships?', *Journal of Applied Pyschology*, Vol. 81, No. 6, pp 795-805

Narayanan L, Menon S, Spector P (1999a), 'A Cross-Cultural Comparison of Job Stressors and Reactions Among Employees Holding Comparable Jobs in Two Countries', *International Journal of Stress Management,* Vol. 6, No. 3

Narayanan L, Menon, S, Spector P (1999b), 'Stress in the Workplace: A Comparison of Gender and Occupations', *Journal of Organizational Behavior,* Vol. 20, pp 63-73

National Institute for Occupational Safety and Health (1997), *NIOSH Generic Job Stress Questionnaire*, Cincinnati, NIOSH

* Neale A V (1991), 'Work Stress in Emergency Medical Technicians', *Journal of Occupational Medicine*, Vol. 33, No. 9

Nelson D, Basu R, Purdie R (1998), 'An Examination of Exchange Quality and Work Stressors in Leader-Follower Dyads', *International Journal of Stress Management,* Vol. 5, No. 2

* Netemeyer R G, Burton S, Johnston M W (1995), 'A Nested Comparison of Four Models of the Consequences of Role Perception Variables', *Organizational Behavior and Human Decision Processes*, Vol. 61, No. 1, pp 77-93

* Netemeyer R G, Johnston M W, Burton S (1990), 'Analysis of Role Conflict and Role Ambiguity in a Structural Equations Framework', *Journal of Applied Psychology*, Vol. 75, No. 2, pp 148-157

Newton T, Keenan A (1985), 'Coping with work-related stress', *Human Relations*, 38, 107-126

Newton T J, Keenan A (1990), 'The Moderating Effect of the Type A Behaviour Pattern and Locus of Control upon the Relationship Between Change in Job Demands and Change in Psychological Strain', *Human Relations,* Vol. 43, No. 12, pp 1229-1255

Niedhammer I, Bugel I, Goldberg M, Leclerc A, Guerguen A (1998), 'Psychosocial Factors at Work and Sickness Absence in the Gazel Cohort: A Prospective Study', *Occupational Environmental Medicine*, Vol 55, pp 735-741

Noboru I, Suzuki K, Kazuo Saito M, Kazuhiko A (1992), 'Type A Personality, Work Stress and Psychologial Distress in Japanese Adult Employees', *Stress Medicine*, Vol. 8, pp 11-21

Norusis M J (1988), 'SPSS-X Advanced Statistics Guide', 2nd ed., *Chicago: SPSS*

Nunnally J (1978), 'Psychometric Theory', 2nd ed., *New York: McGraw Hill*

* O'Driscoll M P, Beehr T A (1994), 'Supervisor Behaviors, Role Stressors and Uncertainty as Predictors of Personal Outcomes for Subordinates', *Journal of Organizational Behavior*, Vol. 15, pp 141-155

Oppenheim A N (1992), 'Questionnaire Design, Interviewing and Attitude Measurement', *London: St Martin's Press*

Orth-Gomer K, Schneiderman N (1996), 'Behavioural Medicine Approaches to Cardiovascular Disease Prevention', *Lawrence Erlbaum Associates*

Osipow S, Spokane A (1987), 'Manual for Occupational Stress Inventory: Research Version', *Odessa FL: Psychological Assessment Resources*

Osipow S H, Davis A S (1988), 'The Relationship of Coping Resources to Occupational Stress and Strain', *Journal of Vocational Behaviour,* Vol. 32, p 1-15

* Parker S K, Sprigg C A (1999), 'Minimizing Strain and Maximising Learning: The Role of Job Demands, Job Control, and Proactive Personality', *Journal of Applied Psychology*, Vol. 84, No. 6, pp 925-939

* Peter R, Siegrist J (1997), 'Chronic Work Stress, Sickness Absence, and Hypertension in Middle Managers: General or Specific Sociological Explanations?', *Soc. Sci. Med.*, Vol. 45, No. 7, pp 111-1120

* Peter R, Geissler H, Siegrist J (1998), 'Associations of Effort-Reward Imbalance at Work and Reported Symptoms in Different Groups of Male and Female Public Transport Workers', *Stress Medicine*, Vol. 14, pp 175-182

Peter R, Siegrist J (1997), 'Chronic Work Stress, Sickness Absence, and Hypertension in Middle Managers: General or Specific Sociological Explanations?', *Social Science and Medicine*, Vol. 45, No. 7, pp 111-1120

* Peters L H, O'Connor E J (1980), 'Situational Constraints and Work Outcomes: The Influences of a Frequently Overlooked Construct', *Academy of Management Review*, Vol. 5, No. 3, pp 391-397

* Phelan J, Bromet E J, Schwartz J E, Dew M A, Curtis E C (1993), 'The Work Environments of Male and Female Professionals, Objective and Subjective Characteristics', *Work and Occupations*, Vol. 20, No. 1, pp 68-89

* Piltch C A, Chapman Walsh D, Mangione T W, Jennings S E (1976), 'Gender, Work and Mental Distress in an Industrial Labor Force: An Expansion of Karasek's Job Strain Model', *American Psychological Association*

Quick J C, Quick J D, Nelson D L and Hurrell J J (1997), *Preventive stress management in organisations*, APA

Quinn R P, Shepard L J (1974), *The 1972-1973 Quality of Employment Survey*, Ann Arbor, MI, Survey Research Centre

* Rees D W, Cooper C L (1991), 'A Criterion Oriented Validation Study of the OSI Outcome Measures on a Sample of Health Service Employees', *Stress Medicine*, Vol. 7, pp 125-127

* Rees D W, Cooper C L (1992), 'The Occupational Stress Indicator Locus of Control Scale: should this be regarded as a state rather than trait measure?', *Work and Stress*, Vol. 6, No. 1, pp 45-48

* Rees D, Cooper C L (1992), 'Occupational Stress in Health Service Workers in the UK', *Stress Medicine*, Vol. 8, pp 79-90

* Renault de Moraes L F, Swan J A, Cooper C L (1993), 'A Study of Occupational Stress Among Government White-Collar Workers in Brazil Using the Occupational Stress Indicator', *Stress Medicine*, Vol. 9, pp 91-104

Revicki D A, May H J, Whitley T W (1991), 'Reliability and Validity of the Work-Related Strain Inventory Among Health Professionals', *Behavioral Medicine,* Vol. 17, No. 3

Rick J, Hillage J, Honey S, Perryman S (1997), 'Stress: Big Issue, but What are the Problems', IES

Rick J, Briner R (2000), 'Psychosocial Risk Assessment: Problems and Prospects', *Occupational Medicine*, Vol. 50, No. 5, pp 310-314

Rick J, Young K, Guppy A (1998), 'From Accidents to Assaults. How organisational responses to traumatic incidents can prevent Post-traumatic Stress Disorder (PTSD) in the workplace, HSE

* Rizzo J R, House R J, S I Lirtzman (1970), 'Role Conflict and Ambiguity in Complex Organizations', *Administrative Science Quarterly*, Vol. 15, pp 150-163

Roberts R, Hemingway H, Marmot M (1997), 'Psychometric and Clinical Validity of the SF-36 General Health Survey in the Whitehall II Study', *British Journal of Health Psychology*, Vol. 2, pp 285-300

* Robertson I T, Cooper C L, Williams J (1990), 'The Validity of the Occupational Stress Indicator', *Work and Stress*, Vol. 4, No. 1 pp 29-39

* Rout U, Cooper C L, Rout J K (1996), 'Job Stress Among British General Practitioners: Predictors of Job Dissatisfaction and Mental Ill-Health', *Stress Medicine*, Vol. 12, pp 155-166

* Russinova V, Vassileva L, Randev P, Jiliova S, Cooper C L (1997), 'Psychometric Analysis of the First Bulgarian Version of the Occupational Stress Indicator (OSI)', *International Journal of Stress Management*, Vol. 4, No. 2

Sawyer J E (1992), 'Goal and Process Clarity: Specification of Multiple Constructs of Role Ambiguity and a Structural Equation Model of Their Antecedents and Consequences', Vol. 77, No. 2, p 130-142

* Schaubroeck J, Ganster D C, Kemmerer B (1996), 'Does Trait Affect Promote Job Attitude Stability?', *Journal of Organizational Behavior*, Vol. 17, pp 191-196

Schaubroeck J, Merritt D E (1997), 'Divergent Effects of Job Control on Coping with Work Stressors: The Key Role of Self-Efficacy', *Academy of Management Journal,* Vol. 40, no. 3, p 738-754

* Schreurs P J, Taris T W (1998), 'Construct Validity of the Demand-Control Model: A Double Cross-Validation Approach', *Work and Stress*, Vol. 12, No. 1, pp 66-84

Schriesheim C A, Eisenbach R J (1995), 'An Exploratory and Confirmatory Factor-Analysis Investigation of Item Wording Effects on the Obtained Factor Structures of Survey Questionnaire Measures', *Journal of Management*, Vol. 21, No. 6 pp 1177-1193

* Schriesheim C A, R J Eisenbach (1995), An Explanatory and Confirmatory Factor-Analytic Investigation of Item Wording Effects on the Obtained Factor Structures of Survey Questionnaire Measures, Journal of Management, Vol. 21, No. 6, pp 1177-1193

* Schuler R S, Aldag R J, Brief A P (1977), 'Role Conflict and Ambiguity: A Scale Analysis', *Organizational Behavior and Human Performance*, Vol. 20, pp 111-128

Searle B J, Bright J E H, Bochner S (1999), 'Testing the 3-factor Model of Occupational Stress: The Impact of Demands, Control and Social Support on a Mail Sorting Desk', Vol. 13, No. 3, pp 268-279

Sen M (1992), 'Retrospected and Anticipated Fits: An Exploration into their Differential Effects in a Sample of Indian Managers', *Work & Stress,* Vol. 6, No. 2, pp 153-162

* Setterlind S, Larsson G (1995), 'The Stress Profile: A Psychosocial Approach to Measuring Stress', *Stress Medicine*, 11, pp 85-92

Shrout P E, Fleiss J L (1979), 'Intraclass correlations: uses in assessing rater reliability', *Psychological Bulletin*, 86, 420-428

Siegrist J, Siegrist K, Ingbert W (1986), 'Sociological Concepts in the Etiology of Chronic Disease: The Case of Ischemic Heart Disease', *Soc. Sci. Med.,* Vol. 22, No. 2, pp 247-253

Siegrist J, Klein D (1990), 'Occupational Stress and Cardiovascular Reactivity in Blue-Collar Workers', *Work & Stress,* Vol. 4, No. 4, pp 295-304

* Siegrist J (1990), (ed. Cooper C L), 'Adverse Health Effects of Effort-Reward Imbalance at Work: Theory, Empirical Support, and Implications for Prevention, Theories of Organizational Stress', *Oxford University Press*, Chapter 9, pp 190-204

* Siegrist J (1996), 'Adverse Health Effects of High-Effort/Low-Reward Conditions', *Journal of Occupational Health Psychology*, Vol. 1, No. 1, pp 27-41

Siegrist J, Peter R (1994), 'Job Stressors and Coping Characteristics in Work-Related Disease: Issues of Validity', *Work and Stress*, Vol. 8, No. 2, pp 130-140

* Siegrist J, Peter R, Junge A, Cremer P, Seidel D (1990), 'Low Status Control, High Effort at Work and Ischemic Heart Diseaase: Prospective Evidence from Blue-Collar Men', *Social Science and Medicine*, Vol. 31, No. 10 pp 1127-1134

Siegrist J and Peter R (1996), *Measuring Effort-Reward Imbalance at Work. Guidelines*, University of Dusseldorf

* Sims J (1995), 'Individual Differences in the Perception of Occupational Stress and their Association with Blood Pressure Status', *Occupational Stress and Blood Pressure Status*, Vol. 9, No. 4, pp 502-512

* Siu O , Donald I, Cooper C L (1997), 'The Use of the Occupational Stress Indicator (OSI) in Factory Workers in China', *International Journal of Stress Management*, Vol. 4, No. 3, pp 171-182

* Siu O, Cooper C L (1998), 'A Study of Occupational Stress, Job Satisfaction and Quitting Intention in Hong Kong Firms: The Role of Locus of Control and Organizational Commitment', *Stress Medicine*, Vol. 14, pp 55-66

* Siu O, Cooper C L, Donald I (1997), 'Occupational Stress, Job Satisfaction and Mental Health Among Employees of an Acquired TV Company in Hong Kong', *Stress Medicine*, Vol. 13, pp 99-107

Slutter J K, van der Beek A J, Frings-Dresen M H W (1999), 'The Influence of Work Characteristics on the Need for Recovery and Experienced Health: A Study on Coach Drivers', *Ergonomics,* Vol. 42, No. 4, p 573-583

* Smith C S, Tisak J (1993), 'Discrepancy Measures of Role Stress Revisited. New Perspectives on Old Issues', *Organizational Behavior and Human Decision Processes*, Vol. 56, pp 285-307

* Smith C S, Tisak J, Hahn S E, Schmieder R A (1997), 'The Measurement of Job Control', *Journal of Organizational Behavior*, Vol. 18, pp 225-237

* Smith C S, Tisak J, Schmieder R A (1993), 'The Measurement Properties and the Role Conflict and Role Ambiguity Scales: A Review and Extension of the Empirical Research', *Journal of Organizational Behavior*, Vol. 14, pp 37-48

* Smulders P G, Nijhuis F J (1999), 'The Job Demands — Job Control Model and Absence Behaviour: results of a 3-year Longitudinal Study', *Work and Stress*, Vol. 13, No. 2, pp 115-131

Soderfeldt B, Soderfeldt M, Muntaner C, O'Campo P, Warg L, Ohlson C (1996), 'Psychosocial Work Environment in Human Service Organisations: A Conceptual Analysis and Development of the Demand-Control Model', *Social Science and Medecine,* Vol. 42, No. 9, pp 1217-1226

* Sparks K, Cooper C L (1999), 'Occupational Differences in the Work-Strain Relationship: Towards the Use of Situation-Specific Models', *Journal of Occupational and Organizational Psychology*, Vol. 72, pp 219-229

* Spector P E (1998) (ed. Cooper C L), 'A Control Theory of the Job Stress Process, Theories of Organizational Stress', *Oxford University Press*, Chapter 7, pp 153-169

* Spector P E (1987), 'Interactive Effects and Perceived Control and Job Stressors on Affective Reactions and Health Outcomes for Clerical Workers', *Work and Stress*, Vol. 1, No. 2, pp 155-162

* Spector P E, Jex S M (1998), 'Development of Four Self-Report Measures of Job Stressors and Strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory and Physical Symptoms Inventory', *Journal of Occupational Health Psychology*, Vol. 3, No. 4 pp 356-367

* Spector P E, Jex S M (1991), 'Relations of Job Characteristics From Multiple Data Sources With Employee Affect, Absence,

Turnover Intentions, and Health', *Journal of Applied Psychology*, Vol. 76, No. 1, pp 46-53

* Spector P E, Jex S M, Chen P Y (1995), 'Relations of Incumbent Affect-Related Personality Traits with Incumbent and Objective Measures of Characteristics of Jobs', *Journal of Organizational Behavior*, Vol. 16, pp 59-65

* Spector P E, Van Katwyk P T, Brannick M T, Chen P Y (1997), 'When Two Factors Don't Reflect Two Constructs: How Item Characteristics Can Produce Artifactual Factors', *Journal of Management,* Vol. 23, No. 5, pp 659-677

Spector P E (2000), 'Industrial and Organizational Psychology: Research and Practice', 2nd ed. *New York: Wiley*

Spector P E, Zapf D, Chen P Y, Frese M (2000), 'Why negative affectivity should not be controlled in job stress research: don't throw out the baby with the bath water', *Journal of Organizational Behavior*, 21, 79-95

* Spielberger C D, Reheiser E C (1994), 'Job Stress in University, Corporate, and Military Personnel', *International Journal of Stress Management*, Vol. 1, No. 1

* Spielberger C D, Reheiser E C (1995), 'Measuring Occupational Stress: The Job Stress Survey', *Occupational Stress: A Handbook* (eds R Crandall, P Perrewe), pp 51-69

Spielberger C D (1994), *Professional manual for the job stress survey,* Odessa, FL, Psychological Assessment Survey

Stansfeld S A, Davey Smith G, Marmott M (1993), 'Association Between Physical and Psychological Morbidity in the Whitehall II Study', *Journal of Psychosomatic Research,* Vol. 37, No. 3, pp 227-238

* Stansfeld S A, Fuhrer R, Head J, Ferrie J, Shipley M (1997), 'Work and Psychiatric Disorder in the Whitehall II Study', *Journal of Psychosomatic Research*, Vol. 43, No. 1 pp 73-81

* Stansfeld S A, Fuhrer R, Shipley M J, Marmot M G (1999), 'Work Characteristics Predict Psychiatric Disorder: Prospective Results from the Whitehall II Study', *Occupational and Environmental Medicine*, Vol. 56, pp 302-307

* Stansfield S A, Bosma H, Hemingway H, Marmot M G (1998), 'Psychosocial Work Characteristics and Social Support as Predictors of SF-36 Health Functioning: The Whitehall II Study', *Psychosomatic Medicine*, Vol. 60, pp 247-255

Stansfeld S, Marmot M (1992), 'Deriving a Survey Measure of Social Support: The Reliability and Validity of the Close

Persons Questionnaire', *Soc. Sci. Med.,* Vol. 35, No. 8, pp 1027-1035

* Stansfield S A, North F M, White I, Marmot M G (1995), 'Work Characteristics and Psychiatric Disorder in Civil Servants in London', *Journal of Epidemiology and Community Health*, Vol. 49, pp 48-53

* Sutherland L F, Fogarty G J (1995), 'Congruence as a Predictor of Occupational Stress', *Journal of Vocational Behaviour*, Vol. 46, pp 292-309

* Sutherland V J, Cooper C L (1993), 'Identifying Distress Among General Practitioners: Predictors of Psychological Ill-Health and Job Dissatisfaction', *Social Science Medicine*, Vol. 37, No. 5, pp. 575-581

* Sutherland V J, Cooper C L (1996), 'Stress in the Offshore Oil and Gas Exploration and Production Industries: An Organizational Approach to Stress Control', *Stress Medicine*, Vol. 12, pp 27-34

* Sutherland V J, Davidson M J (1993), 'Using a Stress Audit: The Construction Site Manager Experience in the UK', *Work and Stress*, Vol. 7, No. 3, pp 273,286

* Swanson V, Power K, Simpson R (1996), 'A Comparison of Stress and Job Satisfaction in Female and Male GPs and Consultants', *Stress Medicine*, Vol. 12, pp 17-26

Tabachnik B G, Fidell L S (1989), 'Using Multivariate Statistics', 2nd ed., *New York: Harper Collins*

* Taber T D, Taylor E (1990), 'A Review and Evaluation of the Psychometric Properties of the Job Diagnostic Survey', *Personnel Psychology*, Vol. 43 (3), pp 467-500

Theorell T, Tsutsumi A, Hallquist J, Reuterwall C, Hogstedt C, Fredlund, P, Emlund N, Johnson J V (1998), 'Decision Latitude, Job Strain, and Myocardial Infarction: A Study of Working Men in Stockholm', *American Journal of Public Health,* Vol. 88, No. 3

* Theorell T (1998), 'Job Characteristics in a Theoretical and Practical Health Context, Theories of Organizational Stress', *Oxford University Press*, Chapter 10, pp 205-219

* Travers C J, Cooper C L (1993), 'Mental Health, Job Satisfaction and Occupational Stress Among UK Teachers', *Work and Stress*, Vol. 7, No. 3, pp 203-219

* Tubre T C, Collins J M (2000), 'Jackson and Schuler (1995) Revisited: A Meta-Analysis of the Relationships Between

Role Ambiguity, Role Conflict, and Job Performance', *Journal of Management*, Vol. 26, No. 1, pp 155-169

* Turnage J J, Spielberger C D (1991), 'Job Stress in Managers, Professionals and Clerical Workers', *Work and Stress*, Vol. 5, No. 3, pp 165-176

* Turnipseed D L (1994), 'An Analysis of the Influence of Work Environment Variables and Moderators on the Burnout Syndrome', *Journal of Applied Social Psychology*, Vol. 24, No. 9, pp 782-800

* Vagg P R, Spielberger C D (1998), 'Occupational Stress: Measuring Job Pressure and Organizational Support in the Workplace', *Journal of Occupational Psychology*, Vol. 3, No. 4, pp 294-305

* Vagg P R, Spielberger C D (1999), 'The Job Stress Survey: Assessing Perceived Severity and Frequency of Occurence of Generic Sources of Stress in the Workplace', *Journal of Occupational Health Psychology*, Vol. 4, No. 3, pp 288-292,

Vagg P R, Spielberger C D (1999), 'The Job Stress Survey: Assessing Perceived Severity and Frequency of Occurrence of Generic Sources of Stress in the Workplace', *Journal of Occupational Health Psychology*, Vol. 4, No. 3, pp 288-292

Vahtera J, Kivimaki M, Uutela A, Pentti J (2000), 'Hostility and ill health: role of psychosocial resources in two contexts of working life', *Journal of Psychosomatic Research*, Vol. 48, pp 89-98

* Valentine S (1999), 'Assessing Organizational Behavior Models: A Comparison of Linear and Nonlinear Methods', *Journal of Applied Social Pyschology*, Vol. 29, No. 5, pp 1028-1044

* Van Der Doef M, Maes S (1999), 'The Job Demand-Control (-Support) Model and Psychological Well-Being: A Review of 20 Years of Empirical Research', *Work and Stress*, Vol. 13, No. 2, pp 87-114

Wall L T (1999), 'Auditing Stress', *Occupational Medicine,* Vol. 49, No. 5, pp 343-344

Wall T D, Jackson P R, Davids K (1992), 'Operator Work Design and Robotics System Performance: A Serendipitous Field Study', *Journal of Applied Psychology,* Vol. 77, No. 3, pp 353-362

* Wall T D, Jackson P R, Mullarkey S (1995), 'Further Evidence on Some New Measures of Job Control, Cognitive Demand and Production Responsibility', *Journal of Occupational and Behavior*, Vol. 16, pp 431-455

* Wall T D, Jackson P R, Mullarkey S, Parker S K (1996), 'The Demands-Control Model of Job Strain: A More Specific Test', *Journal of Occupational and Organizational Psychology*, Vol. 69, pp 153-166

Wall T D, Bolden R I, Borrill C S, Carter A J, Golya D A, Hardy G E, Haynes C E, Rick J E, Shapiro D A, West M A (1997), 'Minor Psychiatric Disorder in NHS Trust Staff: Occupational and Gender Differences', *British Journal of Psychiatry,* 171, pp 519-523

* Warr P B (1990), 'Decision Latitude, Job Demands and Employee Well-Being', *Work and Stress*, Vol. 4, No. 4, pp 285-294

* Westman M (1992), 'Moderating Effect of Decision Latitude on Stress-Strain Relationships: Does Organizational Level Matter?', *Journal of Organizational Behavior*, Vol. 13, pp 713-722

* Westman M (1992), 'Moderating Effect of Decision Latitude on Stress-Strain Relationship: Does Organizational Level Matter?', *Journal of Organizational Behavior*, Vol. 13, pp 713-722

Wheatley D (1990), 'The stress profile', *British Journal of Psychiatry*, 156, pp 685-688

* Williams C M (1997), *Occupational Stress in a National Health Service Trust: An Analysis of the Theoretical Model of the Occupational Stress Indicator*, Thesis for Doctor of Clinical Psychology, Leeds University, School of Medicine

* Williams J S (1996), 'A Critical Review and Further Development of the Occupational Stress Indicator', *Abstract of Thesis for PhD*, UMIST

Williams L J, Gavin M B, Williams M L (1996), 'Measurement and Non-measurement Processes With Negative Affectivity and Employee Attitudes', Vol. 81, No. 1, pp 88-101

* Williams A, Cooper C (1998), 'Measuring Occupational Stress: Development of the Pressure Management Indicator', *Journal of Occupational Health Psychology*, 3, 306-321

* Xie J L (1996), 'Karasek's Model in the People's Republic of China: Effects of Job Demands, Control, and Individual Differences', *Academy of Management Journal*, Vol. 39, No. 6, pp 1594-1618

* Yousef D A (1999), 'Antecedents and Consequences of Job Stressors: A Study in a Third World Setting', *International Journal of Stress Management*, Vol. 6, No. 4, pp 265-282

* Zellars K L , Perrewe P L, Hochwarter W A (1999), 'Mitigating Burnout Among High-NA Employees in Health Care: What Can Organizations Do?', *Journal of Applied Social Psychology*, Vol. 29, No. 11, pp 2250-2271

Zohar D (1995), 'The Justice Perspective of Job Stress', *Journal of Organizational Behavior,* Vol. 16, pp 487-495

* Zohar D (1997), 'Predicting burnout with a hassles based measure of role demands', *Journal of Organizational Behavior*, 18, 101-115

Zohar D (1999), 'When Things Go Wrong: The Effect of Daily Work Hassles on Effort, Exertion and Negative Mood', *Journal of Occupational and Organizational Psychology*, Vol. 72, pp 265-283

* An asterisk denotes papers reviewed.

Printed and published by the Health and Safety Executive
C1 07/01

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from booksellers

HEALTH AND SAFETY INFORMATION
HSE InfoLine
Tel: 08701 545500
Fax: 02920 859260
e-mail: hseinformationservices@natbrit.com
or write to:
HSE Information Services
Caerphilly Business Park
Caerphilly CF83 3GG

HSE website: www.hse.gov.uk

CRR 356

£20.00

ISBN 0-7176-2064-6

9 780717 620647