Journal of Occupational Health Psychology
2000, Vol 5, No 1, 142–155

1076-8998/00/$5.00 DOI: 10 1037//1076-8998.5.1.142

The Validity and Reliability of the Dutch Effort–Reward Imbalance Questionnaire

Eamonn K. S. Hanson
Utrecht University

Wilmar Schaufeli
Utrecht University and The Netherlands Organisation for Scientific Research

Tanja Vrijkotte and Nico H. Plomp
Free University

Guido L. R. Godaert
Utrecht University

The reliability and validity of the Effort–Reward Imbalance Questionnaire were tested in 775 blue- and white-collar workers in the Netherlands. Cronbach's alpha revealed sufficient internal consistency of all subscales except Need for Control. With exploratory probabilistic scaling (Mokken) analysis, the psychometric qualities of the Need for Control scale were improved. With confirmatory factor analysis, the factorial validity of the Extrinsic Effort and Reward subscales was confirmed. A model with 3 separate dimensions for reward (status control, esteem reward, and monetary gratification) proved adequate, emphasizing the importance of distinguishing subscales. The congruent validity of the subscales and a hypothesized relationship with an external construct, health functioning, were confirmed.

The impact of the work environment on health and well-being has been amply demonstrated (e.g., Cooper & Payne, 1991; Hackman & Oldham, 1980; Kahn, 1981; Karasek & Theorell, 1990; Kasl, 1996; Marmot, 1994; Parkes, 1994; Sauter, Hurrell, & Cooper, 1989; Schnall, Landsbergis, & Becker, 1994; Warr, 1987, 1994). Contemporary theories about the adverse effects of the work environment are strongly influenced by three theoretical approaches: the social exchange theory (Adams, 1963; Homans, 1961), the person–environment fit approach (Caplan & Jones, 1975; French, Caplan, & Harrison, 1982; French & Kahn, 1962), and the demand–control approach of Karasek (1979). More recently, the effort–reward imbalance theory (Siegrist, 1996b) has been developed concerning the relationship between the work environment and employee health. A researcher wanting to study the effects of this theory is confronted with the problem of how to measure effort–reward imbalance. Siegrist (1996a) used both subjective and more objective methods to measure imbalance. Subjective methods (i.e., perceptions assessed by paper and pencil) have been associated with risk of coronary heart disease (Bosma, Peter, Siegrist, & Marmot, 1998), but also with health functioning (Stansfeld, Bosma, Hemingway, & Marmot, 1998). On the basis of these findings, and on the fact that perceptual measures have many advantages (e.g., replication of results in different studies; see also Zapf, Dormann, & Frese, 1996), we chose this assessment method. In the present article, we investigate the reliability and validity of a questionnaire to assess extrinsic effort, reward, and need for control (three central constructs in the effort–reward imbalance theory). These analyses will increase the psychometric quality of the Effort–Reward Imbalance Questionnaire, hence adding to the confidence in this measure. First, as a background for this study, the theory of effort–reward imbalance is introduced.

As mentioned, one of the most influential approaches of organizational stress is the *person–environment fit* approach (French & Kahn, 1962). This approach assumes that a discrepancy between environmental demands and an individual's capabilities may lead to mental and physical stress reactions. However, the approach has two main shortcomings. First, it does not specify exactly which of the several objective and subjective aspects of the environment

Eamonn K. S. Hanson and Guido L R. Godaert, Department of Clinical and Health Psychology, Utrecht University, Utrecht, the Netherlands, Wilmar Schaufeli, Department of Social and Organisational Psychology, Utrecht University, and the Netherlands Organisation for Scientific Research, The Hague, the Netherlands; Tanja Vrijkotte, Department of Psychophysiology, Free University, Amsterdam, the Netherlands; Nico H. Plomp, Department of Social Medicine, Free University

We thank Madelon Peters and Miriam Valk-Draad, who assisted the translation of the Effort–Reward Imbalance Questionnaire, and Dirk Enzmann for his assistance with the data analysis.

Correspondence concerning this article should be addressed to Eamonn K. S. Hanson, National Aerospace Laboratory NLR, Man-Machine Integration and Human Factors Research, P O. Box 90502, 1006 BM Amsterdam, The Netherlands. Electronic mail may be sent to hanson@nlr nl.

are responsible for the mental and physical stress reactions. Second, the approach does not explain why a participant does not adapt to the work environment, when a person–environment misfit is experienced.

The first shortcoming has been addressed by Karasek (1979; Karasek & Theorell, 1990) in the *demand–control* approach. Karasek stated that a comprehensive analysis of the work environment entails two important elements of work: (a) *job demands* or workload and (b) *decision latitude* or control over objective task characteristics. According to the approach, stress reactions (e.g., distress, sickness absence, and depression) and associated physiological states (e.g., increased blood pressure) are caused by an interaction between high job demands and poor decision latitude (Marmot & Theorell, 1988). Since the introduction of the demand–control model, it has been extensively tested and has been proven to predict cardiovascular disease (Schnall et al., 1994) in particular. However, the approach has not remained entirely free of critique. First, it focuses on job demands as being the most relevant stress component, underestimating the role played by sociological factors. Second, it does not take individual differences (as buffers or enhancers of the stressor–strain relation) into account.

Social exchange theories (such as the *effort–reward* imbalance theory) offer an explanation for the prevalence of a situation (person–environment mismatch) that seems nonbeneficial to the participant. These theories emphasize that it is the perception and evaluation of social exchange in relationships (i.e., between individual and organization) that determine successful functioning and health. Inequitable situations lead to distress, which motivates individuals to restore equity (Walster, Berscheid, & Walster, 1973). It is assumed that individuals will strive to maximize their outcomes and minimize their inputs (Adams, 1965). Having a certain personality (e.g., Type A) may add to this tendency.

One of the most likely places for effort–reward imbalance to occur is in the work environment. Perceived imbalance will occur if the extrinsic effort (i.e., time pressure, increasing demands, and responsibility) that is spent during work does not correspond with the reward that is obtained. As a result, the employee feels that his or her social status is threatened, leading to emotional distress, changes in physiological (re)activity, and eventually to cardiovascular disease. The term *extrinsic effort* refers to the perceived working conditions For instance, an employee may have the impression that she or he always works under a high time pressure because of a high workload or that he or she is frequently interrupted during work.

According to the effort–reward imbalance theory, three types of "reward" are distinguished: (a) status control, (b) esteem reward, and (c) monetary gratification (Matschinger, Siegrist, Siegrist, & Dittman, 1986). *Status control* refers to the opportunity a person has to achieve, or have control over, a desired social status (e.g., promotion prospects, job security, and status inconsistency). In contrast to control in the demand–control approach, status control reflects the influence of fragmented job careers, job instability, redundancy, and forced occupational downward mobility. Under these conditions, concerns about or even the benefits of task control may be overridden. *Esteem reward* refers to experienced respect and support during work, and *monetary gratification* refers to the job salary in relation to coworkers.

In addition, a construct was introduced to describe the persistence of a situation (person–environment mismatch) that seems nonbeneficial to the participant: *need for control* (which consists of need for approval, competitiveness, disproportionate irritability, and inability to withdraw from work). This construct is closely related to aspects of the Type A behavior pattern (e.g., hostility) that are linked to enhanced arousal in demanding situations. Participants with a high need for control also tend to misjudge (i.e., overestimate or underestimate) demanding situations. Both enhanced arousal and misjudging the situation are considered instrumental in eliciting excessive efforts (even in situations that seem nonbeneficial to the participant). Need for control may therefore be considered as a generalized coping strategy, instrumental in eliciting excessive efforts that may last for considerable lengths of time and probably linked with activation of the autonomic nervous system. Research shows that the significance of the relation between effort–reward imbalance and cardiovascular disease is increased (after controlling for traditional risk factor such as age, body mass index, blood pressure, and low-density lipoprotein cholesterol), if individuals respond to the situation with a high need for control (Siegrist & Peter, 1994).

The aim of the present study was to assess and, if necessary, improve the reliability and validity of the Effort–Reward Imbalance Questionnaire. We achieved this by examining the reliability, factorial validity, congruent validity, and content validity of extrinsic effort, reward, and need for control in a Dutch population. We determined reliability by assessing the internal consistency (i.e., by calculating reliability coefficients or Cronbach's alpha) for each of the subscales of the questionnaire. We determined factorial validity by testing whether the items loaded on the factors described in the theory using a first-order confirmatory factor analysis (CFA). We tested congruent validity by assessing the fit of a model in which all Effort–Reward Imbalance subscales load on the same second-order factor (or latent variable). Finally, we assessed content validity by determining its relationship with an external reference: health functioning.

Health functioning was measured using a questionnaire (the Medical Outcomes Survey Short Form; MOS SF-20) that covers three dimensions of health: social, mental, and physical (Veit & Ware, 1983). The MOS, originally developed to assess the outcomes of medical care, is the most widely used self-report measure of functioning (Stansfeld et al., 1998) and has been found to be a reliable and valid instrument in population studies. Basically, it is hypothesized that the scales measuring extrinsic effort and need for control as well as the latent variable, reward, could be distinguished from the subscales that measure mere self-reported health. Therefore, a model is tested in which no covariance between these (sub)scales is allowed. However, in the literature, it has been established that work stress was associated with negative health outcomes (Stansfeld et al., 1998). Therefore, we tested a second model in which covariance was allowed between the second-order factors work stress and health functioning but not between the first-order factors (extrinsic effort, status control, esteem reward, and need for control on the one hand, with physical functioning, role functioning, social functioning, mental health, health perceptions, and pain on the other) Before the MOS SF-20 was used as an external reference, its congruent validity was determined.

Method

Participants

In the present study, 775 employees (*M* age = 43 years, *SD* = 6 62) from four companies returned questionnaires that addressed work- and health-related issues. The sample consisted of both blue- and white-collar workers from different work settings. software specialists (60%), national railway personnel (25%), health professionals (9%), and office clerks (6%) Of the participants, 82% were male and 18% female

Procedure

Before statistical analyses were carried out, the sample was divided into two equally sized subsamples· test construction (n = 367) and validation (n = 369). These subsamples were matched according to age, sex, educational background, and company. Statistical analyses were performed in two steps. In the first step, the test-construction subsample was used to assess construct validation and to eventually develop new (sub)scales using exploratory statistical techniques. In the second step, the validation subsample was used to confirm the psychometric properties of the (sub)scales previously constructed.

Measures

The German Effort–Reward Imbalance Questionnaire ("Zentrale Fragen für die Erfassung von Gratifikationskrisen am Arbeidsplatz," by Siegrist and coworkers; Siegrist, 1996b) was used to measure the employees' effort–reward imbalance. This 47-item questionnaire was translated into Dutch, and the adequacy of the translation was checked by a Dutch-German bilingual social scientist. In addition, the questionnaire was translated back into German by an independent professional translator. Differences that arose after the back-translation into German were compared with the original version, and the questionnaire was adapted accordingly.

The questionnaire measures three main constructs: Extrinsic Effort, Reward, and Need for Control. Extrinsic Effort was measured by six items that refer to demanding aspects of the work environment (e.g., "I have constant time pressure due to a heavy workload"). If the participants answered the question affirmatively, they were then asked to rate the severity of this ranging from *not at all distressed* (1) to *very distressed* (4) A negative answer to the question (indicating the absence of effort or that the question was out of order) also scored 1 point. Reward was measured by 12 items that form two subscales· Esteem Reward (5 items, e.g., "I receive the respect I deserve from my colleagues") and Status Control (6 items, e g., "My promotion prospects are poor"). The last item is referred to as monetary gratification (1 item: "Considering all my efforts and achievements, my salary/income is adequate") The Reward items were scored in the same way as the Extrinsic Effort items, so that a minimum score of 1 point and a maximum score of 4 points per item could be obtained.

The Need for Control scale consists of 29 dichotomous items (*disagree* = 0, *agree* = 1) that form four subscales Need for Approval (6 items, e g., "I only feel successful when I perform better than I expected"), Competitiveness (6 items, e g., "I don't let others do my work"), Disproportionate Irritability (8 items, e.g., "Even the slightest interruption bothers me"), and Inability to Withdraw From Work (9 items, e g , "Work is usually still on my mind when I go to bed"). The internal consistency of the original (German) version of the Reward subscales (Esteem Reward and Status Control) and Need for Control scale is considered satisfactory (see Dittmann & Matschinger, 1982; Matschinger et al , 1986, Siegrist & Peter, 1994) The same goes for the Dutch translation of the Effort, Reward, and revised Need for Control subscales (see Table 1)

Health functioning was measured by administering the Dutch version (Kempen, 1992a) of the MOS SF-20 questionnaire (Stewart, Hays, & Ware, 1988) in a subsample consisting of railway personnel, health professionals, and office clerks (n = 226) The items of the MOS SF-20 provide information about health functioning and quality of life,

Table 1
Descriptive Statistics and Alpha Coefficients for the Effort–Reward Imbalance Questionnaire (Original and Revised Subscales: N = 775) and MOS SF-20 Subscales (N = 226)

Subscale	*n* items	*M*	*SD*	α
Extrinsic Effort	6	10.9	3.0	.71
Reward				
Status control	6	19.9	3.6	.70
Esteem reward	5	17.5	3.1	.77
Monetary gratification	1	3.6	0.8	
Need for Control (intrinsic effort)				
Need for approval	6	3.5	1.4	.43
Competitiveness	6	1.5	1.5	.59
Time pressure	8	2.8	1.7	.54
Inability to withdraw from obligations	9	3.4	2 3	.68
Total scale (Need for Control)	29	10.5	5.4	.82
Revised Need for Control	9	2.4	2.5	.81
Health measures: MOS SF-20				
Physical functioning	6	73.7	27.9	.76
Role functioning	2	66.7	43.2	.81
Social functioning	1	62.0	26.8	
Mental health	5	70.1	20.6	.88
Health perceptions	5	71.6	19.3	.78
Pain	1	32.6	32.5	

Note MOS SF-20 = Medical Outcomes Survey Short Form

clustered in six dimensions· physical functioning (6 items, e g, "The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?"), role functioning (2 items, e g, "During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems [such as feeling depressed or anxious]?"), social functioning (1 item: "During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities [like visiting with friends, relatives, etc]?"), mental health (5 items, e g, "Have you felt calm and peaceful?"), health perceptions (5 items, e g, "In general would you say your health is . ."), and pain (1 item "How much bodily pain have you had during the past 4 weeks?") Scores were transformed into a 100-point scale, with higher scores reflecting better functioning except for the pain dimension, in which a higher score means more perceived pain The internal consistency of all subscales of the Dutch MOS SF-20 is satisfactory (Cronbach's α > 80 for each subscale, Kempen, 1992a, 1992b, Kempen, Brilman, & Ormel, 1993, Moorer & Suurmeijer, 1993). Kempen et al (1993) showed the subscales of the MOS SF-20 to be externally valid by correlating them with external criteria A high positive correlation (r = 78) was found between the MOS SF-20 subscale Mental Health and Positive Well-Being (Well-Being Scale, Veit & Ware, 1983), as well as with Positive Affect (r = 42, Affect Balance Scale, Bradburn, 1969) Kempen (1992b) also found a positive correlation (r = 71) between the MOS SF-20 subscale Physical Functioning and Movement Restriction.

Data Analysis

In the present study, tests of reliability and validity were performed To enable comparison of data across scales and subscales, we used identical methods to assess reliability and validity. Reliability (internal consistency) was determined by calculating Cronbach's alpha. According to criteria proposed by Nunnaly (1978), an alpha higher than .70 indicates satisfactory internal consistency.

Validity was determined by performing CFAs on the subsamples using EQS.[1] To indicate the fit of the (sub)scales to the data, a number of the most common goodness-of-fit indexes were calculated: the Bentler–Bonnet nonnormed fit index (NFI; Bentler & Bonnet, 1980; Tucker & Lewis, 1973), the comparative fit index (CFI; Bentler, 1989), the LISREL adjusted goodness-of-fit index, and the root mean squared residual. The Satorra–Bentler scaled chi-square (Satorra & Bentler, 1988) and the corrected CFI (CFI*)[2] were also calculated because a number of kurtotic items were detected. These latter two indexes correct distributional abnormality A fit of over .90 for CFI and NFI is considered adequate (Byrne, 1994).

As was described in the procedure, statistical analyses were performed on a test-construction and a validation subsample In the test-construction subsample, the fit of the most plausible model based on theoretical considerations was tested (confirmatory analysis) and, if necessary, was improved (exploratory analysis). Improvement was achieved by using the Lagrange multiplier test, which determines

[1] EQS is statistical software for structural equation modeling (Bentler, 1989)

[2] Corrected comparative fit index (CFI*) based on Satorra–Bentler chi-square fit for the null model = 1 − max[($\chi^2_0 - df_0$), 0] max[($\chi^2_k - df_k$), ($\chi^2_0 - df_0$), 0], where df_0 = degrees of freedom for the null model, df_k = degrees of freedom for the hypothesized model, χ^2_0 = chi-square for the null model, and χ^2_k = chi-square for the hypothesized model (Hu & Bentler, 1995)

whether the specification of certain parameters would lead to a better model in a subsequent EQS run. The goodness of fit of the optimal model was then tested in the second population (validation subsample). The model was accepted if the fit in the second subsample was adequate. The achievement of fit in the second model minimized the chance that the fit acquired in the first model (after respecifications) was due to a capitalization on chance factors.

For one of the scales (Need for Control), we used a revised procedure. Because the items of the Need for Control subscales are dichotomous, the revised scale was constructed according to principles of the probabilistic test theory. An example of a statistical test based on the probabilistic test theory[3] is the Mokken analysis (Mokken, 1971; Niemöller, Schuur, & Stokman, 1980), which can be used to evaluate existing scales or to construct new ones. The psychometric qualities of the scale are determined by calculating a goodness of fit or scalability coefficient (H) and a reliability coefficient (ρ). This coefficient can be used to evaluate a set of items thought to be a scale or for constructing a scale from a given pool of items. Mokken and Lewis (1982) proposed the following criteria for scalability: H < .30 = no scale; .30 = H < .40 = weak scale; .40 = H < .50 = medium scale; H = .50 = strong scale. The scalability of whole scales (H) as well as separate items (H(i)) can be judged by this criterion, and a reliability coefficient (ρ) can be calculated for a set of items. Values over .70 indicate a reliable scale (Niemöller et al., 1980). The Mokken analysis was carried out by using a software program called MSP (or Mokken Scale Analysis for Polychotomous Items; Debets & Brouwer, 1989; Sijtsma, Debets, & Molenaar, 1990).

Results

Reliability and Mean Values

As can be seen in Table 1, all scales had a Cronbach's alpha of .70 or higher, except the original Need for Control subscales (see Table 1). Therefore, we decided to construct a revised Need for Control scale using a Mokken analysis.

Construction of Revised Need for Control

The results of the Mokken analysis are reported in Table 2. Three out of the four subscales had low scalabilities, and only one reliable subscale (Inability to Withdraw From Obligations) was found. The total scale (sum of all the items) may be considered reliable (ρ = 0.84), but it has a low scalability (H = 0.21). These results confirm the findings of the tests for internal consistency.

We continued the analysis by conducting an explorative procedure to find subscales with acceptable scalabilities and reliabilities. The following criteria were applied to the scales and separate items: H(i) = 0.30; H = 0.40, *n* items = 3; ρ = 0.70. This procedure led to the construction of only one scale (see Table 3) with a scalability (H) of 0.45 and a reliability (ρ) of 0.84. The quality of this scale was then confirmed by using the validation subsample within which both scalability and reliability were still above the proposed criteria (H = 0.40 and ρ = 0.81). Cronbach's alpha calculated over the entire population was .81, also indicating a satisfactory internal consistency. For the revised (9-item) version of the scale, the Cronbach's alpha was .81, which was slightly lower than the Cronbach's alpha of the 29-item version (.84). This difference is quite acceptable, because based on the Spearman–Brown formula, a reduction from 29 to 9 items should decrease the alpha by .22. We only found a .03 reduction. Thus, the practicality of the scale has been improved at minimal loss of information and reliability.

Table 2
Scalability (H) and Reliability (ρ) of the Four Original Need for Control Subscales and a Total Scale Score, as Provided by the Mokken Test Procedure

Subscale	H	ρ
Need for Approval	0.12	0.39
Competitiveness	0.31	0.66
Disproportionate Irritability	0.22	0.58
Inability to Withdraw From Obligations	0.28	0.73
Total Need for Control scale	0.21	0.84

Note. Test-construction subsample *N* = 367.

Factorial Validity of Extrinsic Effort, Reward, and Need for Control

The factorial validity was assessed by performing a first-order CFA on the test-construction subsample using EQS (see Footnote 1). The fit of the following models was tested (see Figure 1 and Table 4).

Extrinsic effort. We tested whether the six items of extrinsic effort all loaded on a single latent variable. This model (M1) achieved acceptable fit (CFI* = .95) and fitted the data better than the null model ($\Delta\chi^2 < 0.001$).

Reward. The fit of three models was tested: first, a model (M2) in which all items loaded on a single

[3] This model can be viewed as a probabilistic version of Guttman scales analysis for dichotomous items or as a nonparametric approach to item response theory. Another example of an item response model (although parametric) is the Rasch model (Rasch, 1960), often used for the scaling of questionnaires with dichotomous items.

Table 3
Item Numbers, Mean Scores, and Scalability (H(i)) of the Separate Items of the Revised Need for Control Scale

Item	Scale	Content	*M*(i)	H(i)
3	DI	Even the slightest interruption bothers me	0.13	0.46
14	DI	I can get furious if someone doesn't understand me the first time	0.15	0.38
12	COM	I don't let others do my work	0.18	0.36
17	NFA	I only feel successful when I perform better than I expected	0.26	0.38
29	IWO	If I put off doing something that needs to get done today, I'll have trouble sleeping at night	0.27	0.39
24	IWO	Work is usually still on my mind when I go to bed	0.30	0.47
10	IWO	I start thinking about problems as soon as I get up in the morning	0.39	0.43
15	IWO	When I get home, I can easily relax and forget all about work	0.40	0.39
16	IWO	People close to me say I sacrifice too much for my job	0.41	0.40

Note Test construction subsample $N = 367$. DI = Disproportionate Irritability; COM = Competitiveness; NFA = Need for Approval; IWO = Inability to Withdraw From Obligations.

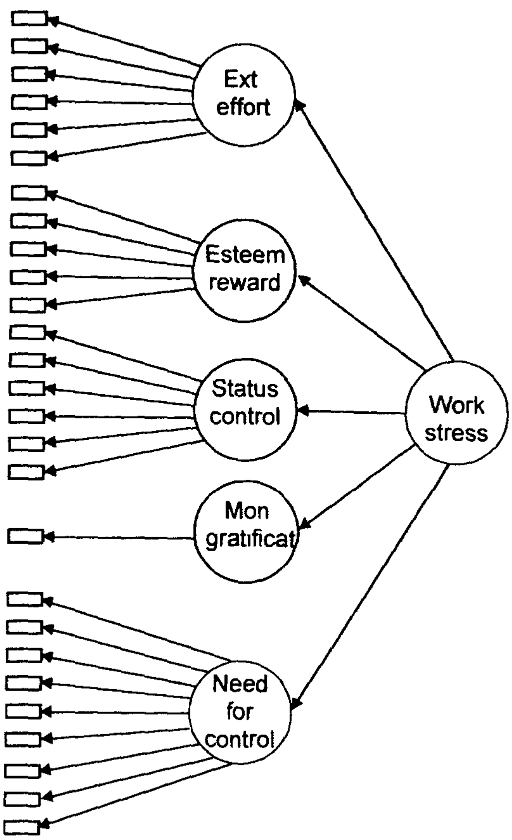

Figure 1 Representation of the model used to test the fit of a model in which extrinsic (Ext.) effort, esteem reward, status control. monetary gratification (Mon gratificat), and need for control load on the second-order factor work stress (see M2 in Table 5) Error terms are not shown in the figure

latent variable, and, second, as hypothesized by Siegrist (1996b), a model (M3) with two latent variables (status control and esteem reward) and a third dimension (monetary gratification). Covariance was allowed between the three dimensions and a third model (M3a), in which three error terms were allowed to correlate. The items that were allowed correlated errors were the following: Item 12 (from the Status Control subscale: "Do you experience or expect an undesirable change in your job situation?"), Item 13 (from the Status Control subscale: "Has job redundancy recently affected your work colleagues?"), and Item 14 (also from the Status Control subscale: "Is your own job security poor?"). Model M3a had the best fit (CFI* = .97) and was significantly better than Model M2 ($\Delta\chi^2 < 0.001$).

Need for control. This model (M4) reflects the nine items of the revised[4] Need for Control scale loading on the same first-order latent variable. The model achieved acceptable fit (CFI* = .92) after two pairs of error terms (E3, E12 and E14, E17; see Table 3 for respective items) were allowed to correlate (M4a) and was significantly better than the null model ($\Delta\chi^2 < 0.001$).

Models M1, M3, and M4 were also tested in the test validation subsample. Again, the fit of Models M3 and M4 was improved by allowing correlated errors (M3a and M4a). The Lagrange multiplier test revealed that the same correlated errors should be allowed for Model M3a (reward) but not for Model M4a (need for control). The test of Model 4a in the test validation subsample achieved adequate fit after

[4] The factor structure of revised need for control has already been tested by means of Mokken analysis. For reasons of comparability with the other scales, a CFA was also performed.

Table 4
Comparison of Factorial Models Representing Extrinsic Effort, Reward, and Need for Control (see Figure 1)

Model	χ^2	*df*	$\Delta\chi^2$	NFI	CFI	AGFI	RMSR	S-B χ^2	CFI*
Test-construction subsample (N = 367)									
M0 null model extrinsic effort	327 85	15						273.34	
M1 extrinsic effort (1 latent variable)	26 86	9	<0.001	.91	94	.94	.026	21.05	.95
M0 null model reward	1,095 19	66						625.90	
M2 reward (1 latent variable)	440 60	54	<0.001	.54	62	75	.151	281.74	.59
M3 reward (3 separate factors)	182 11	52	<0.001	.84	.87	.88	.060	125.37	.87
M3a reward with correlated errors	98.45	49	<0.001	.94	.95	93	.034	67 53	.97
M0 null model need for control	909 69	36						714.70	
M4 need for control (1 latent variable)	163.58	27	<0 001	.79	.84	81	.013	133 89	.84
M4a: with correlated errors	96 58	25	<0.001	.88	.92	.89	012	82 32	.92
Validation subsample (N = 369)									
M0 null model extrinsic effort	464 38	15						343.16	
M1 extrinsic effort (1 latent variable)	23 83	9	<0 001	95	.97	.95	.022	20.53	.97
M0 null model reward	1,385.93	66						775.65	
M3· reward (3 separate factors)	336 26	52	<0.001	.87	.89	.88	.088	267.98	.81
M3a reward with correlated errors	130.64	49	<0.001	.92	.94	.91	.051	91.32	.94
M0· null model need for control	841 41	36						638.17	
M4 need for control (1 latent variable)	203 21	27	<0 001	.71	.78	.76	.017	154.45	.79
M4a· with correlated errors	76 92	24	<0 001	.90	93	.92	.015	61.77	.94

Note. M0 is in each case a test of the null model; M1 tests the factorial validity of extrinsic effort; M2 tests the factorial validity of reward, in which all items load on a single factor; M3 tests the factorial validity of reward with three separate factors (covariance is allowed between the factors); M3a is the same as M3 with correlated errors; M4 tests the factorial validity of need for control, and M4a is the same as M4 with correlated errors The significant change in chi-square of each new model is tested against the preceding model ($\Delta\chi^2$). For all chi-squares, $p <$ 001. NFI = nonnormed fit index; CFI = comparative fit index, AGFI = adjusted goodness-of-fit index; RMSR = root mean square residual; S-B χ^2 = Satorra–Bentler scaled chi-square; CFI* = corrected CFI.

three error terms (E24, E10; E24, E15; and E10, E15; see Table 3) were allowed to correlate. These items refer to the inability to withdraw from obligations, in contrast to the items of the correlated errors in Model 4a.

In summary, the high values of the goodness-of-fit indexes demonstrate adequate factorial validity for extrinsic effort, reward, and need for control. For reward and need for control, fit was reached after allowing correlated errors indicating common variance between the items not reflected by the latent factor. The Lagrange multiplier test in the validation subsample reveals the same correlated errors for reward (M3a) but not for need for control (M4a). The results increase the confidence in the error intercorrelations for reward (M3a) but not for revised need for control (M4a).

Congruent Validity

Congruent validity was tested progressively by CFA in the test-construction subsample. First, a model (M1) was tested in which extrinsic effort, need for control, status control, esteem reward, and monetary gratification formed independent dimensions. The error correlations were derived from previous analyses (see Table 4). Second, in Model M2 all first-order factors were allowed to load on a single second-order factor that can be conceived as work stress (see Figure 1). Model M2a included three new correlations between error terms.[5] Each new model was expected to have a better fit than the preceding model. For instance, the final model (M2a) was expected to have the best fit, followed by Model M2, and so on The goodness-of-fit estimates are shown in Table 5 As anticipated, the chi-square statistic of the null model (M0) was very high, indicating excessive misfit. The fit of the second model (M1) improved significantly, as indicated by ($\Delta\chi^2$). However, only the fit of Model M2a was adequate. The final model (M2a) also had an adequate fit in the validation subsample, hence confirming congruent validity (see Table 5) The Lagrange multiplier test in the validation subsample reveals the same correlated errors for reward (Model M2a), increasing the confidence in the error intercorrelations. In summary, the model with the best fit was the model in which the Effort, Reward, and Need for Control subscales were allowed to load on a single second-order factor (work stress), confirming congruence.

In the introduction, we mentioned that subjectively experienced health was measured in addition to extrinsic effort, reward, and need for control. Subjectively experienced health was used as a second-order factor in the final analysis of this article. Before that final model was tested, we made a three-step assessment (Table 6) to determine the factor structure of subjectively experienced health. First, the null model (M0) was tested. Then, convergence of all factors on one second-order factor was assessed (Model M1). This model obtained an adequate fit after respecification of the errors[6] (Model M1a). All six subscales load on the same second-order factor, health, indicating congruent validity (see Table 6).

Content Validity

To determine content validity, we calculated the Pearson's intercorrelations between the subscales of the Effort–Reward Imbalance Questionnaire. The results are given in Table 7 and show moderate to high intercorrelations for Extrinsic Effort, Status Control, and Esteem Reward, indicating some overlap (in content) between these scales and subscales. Need for Control had a low correlation with the other subscales, which suggests hardly any overlap with the other (sub)scales.

Nearly all subscales of the MOS SF-20 are significantly intercorrelated, except for the correlation between physical functioning and mental health and between mental health and pain. Therefore, it can be concluded that most of the MOS SF-20 subscales are conceptually related.

The correlations between the Effort–Reward Imbalance scales and subscales and those of the MOS SF-20 varied between −.40 and .27, indicating a weak to moderate overlap between a number of the subscales (see Table 7). This is in accordance to what may be theoretically expected. Extrinsic effort, reward, and need for control and subjectively experienced health were distinguishable, but not totally independent constructs. More specifically, a high extrinsic effort was associated with a lower well-being, and in general a high reward was related to higher well-being.

The original Need for Control subscales and the revised scale were both positively correlated with social functioning and negatively correlated with

[5] The errors that were allowed shared variance belong to Items 9 and 10 (from the Esteem Reward subscale), Items 4 (from the Extrinsic Effort subscale) and 6 (from the Need for Control scale), and Items 5 and 2 (from the Extrinsic Effort subscale)

[6] The errors of Items 12 and 10 (mental health) and of Items 20 and 13 (social functioning and mental health) were allowed to correlate

physical functioning, mental health, and health perceptions. Both the original and revised Need for Control scales and subscales correlated highly positive with each other (r = .90). The Pearson's correlations show that extrinsic effort, reward, and need for control are not totally independent.

Finally, the tests of congruent validity were concluded by assessing the fit of two models: (a) a model (M1) with two second-order factors (work stress and health functioning) and no covariance between the two second-order factors, and (b) the same as M1 but allowing for covariance between the second-order factors (Table 8).

The first model did not adequately fit the data. The second model (M2; see Figure 2) in which covariance between the second-order factors was allowed, shows an improvement in fit (p = .039). However the model does not fit the data (goodness-of-fit indexes are below 0.90). Because of the absence of sufficient theoretical support (as to which factors are allowed to cross-load), it did not seem appropriate to respecify the second model to obtain a better fit. The fit of a model that allows covariance between the second-order factors but no covariance between the first-order factors was not confirmed, indicating that both instruments were not totally independent.

On the whole, the Effort–Reward Imbalance Questionnaire and the MOS SF-20 form coherent clusters as is indicated by congruence (congruent validity; see Tables 5 and 6). A model that tested the independence of the effort–reward imbalance as well as the MOS SF-20 factors, although allowing for covariance between the second-order factors, did not fit the data (see Table 8). This indicates that more covariance between the factors (other than between the second-order factors) is probable.

Discussion

In this article, we assessed the reliability, factorial validity, congruent validity, and content validity of the Effort–Reward Imbalance Questionnaire. In accordance with the theory, extrinsic effort and reward (status control, esteem reward, and monetary gratification) were both factorially valid (see Table 4). Yet, a slight weakness was found among the items of the Status Control subscale. More specifically, structural equation modeling revealed three items of this scale to have correlated errors. This means that the items have a shared communality that is absent in the factor they load on. Remarkably, the items with shared error variance all refer to job insecurity (e.g., "Do you experience or expect an undesirable change

Table 5

Determining the Congruent Validity of Effort, Reward, and Revised Need for Control Subscales, by Performing a Second-Order CFA Using the Test-Construction and Validation Subsamples (see Figure 2)

Model	χ^2	*df*	$\Delta\chi^2$	NFI	CFI	AGFI	RMSR	S-B χ^2	CFI*
Test-construction subsample (N = 367)									
M0 null model	2,700.84	351						2,136.36	
M1: five-factor CFA	902.94	319	<0.001	.73	.75	.79	.09	743.62	.76
M2: five-1st-order factors and one 2nd-order factor	604.73	315	<0.001	.86	.88	.86	.03	504.86	.89
M2a (with three correlated errors)	558.83	312	<0.001	.88	.90	.86	.03	467.92	.91
Validation subsample (N = 369)									
M0: null model	2,960.61	351						2,314.02	
M2: five 1st-order factors and one 2nd-order factor	619.42	315	<0.001	.73	.75	.79	.09	743.62	.90
M2a (with three correlated errors)	321.47	312	<0.001	.89	.90	.88	.06	237.46	.91

Note. M0 is in each case a test of the null model, M1 is a test of the independence of extrinsic effort, status control, esteem reward, monetary gratification, and revised need for control (five-factor model, the five factors are not allowed to load on each other); M2 is a test of a model in which extrinsic effort, esteem reward, status control, monetary gratification, and the need for control load on the second-order factor work stress; M2a is the same as M2, except three error terms were allowed to correlate. The significant change in chi-square for each new model is tested against the preceding model ($\Delta\chi^2$). For all chi-squares, p < .001. CFA = confirmatory factor analysis; NFI = nonnormed fit index; CFI = comparative fit index; AGFI = adjusted goodness-of-fit index; RMSR = root mean square residual; S-B χ^2 = Satorra–Bentler scaled chi-square; CFI* = corrected CFI.

Table 6
Determining the Congruent Validity of the MOS SF-20 Subscales by Performing a Second-Order CFA on a Subpopulation

Model	χ^2	*df*	$\Delta\chi^2$	NFI	CFI	AGFI	RMSR	S-B χ^2	CFI*
M0: null model	2.073.86	190						1,806.30	
M1: 2nd-order CFA	432.86	166	<0.001	.84	.86	.78	.09	405.13	.85
M1a	348.91	164	<0.001	.89	.90	.82	.09	326.21	.90

Note. $N = 226$. M0 is a test of the null model; M1 is a test for one underlying latent variable (second-order CFA model); M1a is the same as M1, but it also allows covariance between a number of error terms. The significant change in chi-square of each new model is tested against the preceding model ($\Delta\chi^2$). For all chi-squares, $p < .001$. MOS SF-20 = Medical Outcomes Survey Short Form; CFA = confirmatory factor analysis; NFI = nonnormed fit index; CFI = comparative fit index; AGFI = adjusted goodness-of-fit index; RMSR = root mean square residual; S-B χ^2 = Satorra–Bentler scaled chi-square; CFI* = corrected CFI.

in your work situation?"), whereas the factor they are supposed to load on (status control) entails more than just job security. The importance of these intercorrelations was supported by the test in the validation subsample: The same intercorrelations between error terms improved the fit of the model, which was otherwise inadequate. Therefore, in future studies, it might be fruitful to consider job security as a separate factor of job demands, particularly because future job insecurity has shown to be a major stressor (Hartley, Jacobson, Klandersmans, & van Vuren, 1991).

The original Need for Control subscales revealed low reliabilities and scalabilities. Considering the low reliabilities found in the study reported by Matschinger et al. (1986), these results are not totally surprising. Thus, we concluded that these subscales of Need for Control should not be used in further analyses. Consequently, we constructed a revised Need for Control scale using a Mokken test procedure for dichotomous items. A single scale with an acceptable reliability and scalability was identified. To enable comparisons with the other scales of the Effort–Reward Imbalance scales, we also performed a CFA on the revised Need for Control scale. An acceptable fit was achieved after some error terms were allowed to correlated. The validation subsample revealed other correlated errors than the test-construction subsample. This reduces the confidence in the contribution of these items to the construct Need for Control. Future studies should reassess the factor structure of the revised Need for Control scale to increase the confidence in this scale.

The content (or external) validity of the Need for Control scale (e.g., using more objective measures) also has to be determined in future research. Both the 29-item and the revised (9-item) Need for Control scale are strongly associated with subjectively experienced health. To date, only the 29-item scale has been tested in a longitudinal study to predict future cardiovascular disease. Although we expect the

Table 7
Pearson's Correlation of the Effort–Reward Imbalance and MOS SF-20 Subscales

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1 EE	—											
2 SC	− 36*	—										
3 ER	− 39*	56*	—									
4 MG	− 20*	.39*	39*	—								
5 NfC	.10	.01	− 12*	−.04	—							
6 R-NfC	.13*	−.01	− 13*	− 05	90*	—						
7 PF	− 15*	04	.10	10	−.39*	− 40*	—					
8 RF	− 14*	.10	08	− 05	06	07	.51*	—				
9 SF	− 14*	.16*	14*	− 03	20*	18*	38*	.53*	—			
10 MH	− 20*	24*	27*	00	−.14*	−.13*	08	.21*	.44*	—		
11 HP	− 27*	27	24*	.08	− 02	−.06	34*	22*	31*	38*	—	
12. P	15*	00	− 11	−.03	04	04	−.58*	−.45*	−.32*	−.08	−.35*	—

Note $N = 226$ MOS SF-20 = Medical Outcomes Survey Short Form, EE = extrinsic effort; SC = status control; ER = esteem reward, MG = monetary gratification, NfC = need for control. R-NfC = revised need for control; PF = physical functioning, RF = role functioning, SF = social functioning, MH = mental health; HP = health perceptions; P = pain.
* $p <$ 05, two-tailed

Table 8
Determining (Content) Validity by Performing a CFA on Effort, Reward, and Revised Need for Control and Health Functioning (MOS SF-20)

Model	χ^2	*df*	$\Delta\chi^2$	NFI	CFI	AGFI	RMSR	S-B χ^2	CFI*
M0 null model	5,361.03	1,081						4,717.12	
M1 2nd-order CFA, independent	1,820.81	1,019	<0.001	.80	.81	.71	.08	1,646.03	.83
M2 2nd-order CFA, dependent	1,816.56	1,018	0.039	.80	.81	.71	.08	1,642.81	.83

Note. $N = 226$. M0 is a test of the null model. M1 is a model with two second-order factors (work stress and health). No covariance is allowed between the factors of the Effort–Reward Imbalance Questionnaire (extrinsic effort, status control, esteem reward, monetary gratification, and need for control) and the MOS SF-20 (physical functioning, role functioning, social functioning, mental health, health perceptions, and pain). There is no covariance between the latent factors. Error correlations are the same as in previous models. M2 is the same as M1, only covariance is allowed between the second-order factors. The factors of the Effort–Reward Imbalance Questionnaire and the MOS SF-20 are not allowed to cross-load. The significant change in chi-square of each new model is tested against the preceding model ($\Delta\chi^2$). For all chi-squares, $p < .001$. CFA = confirmatory factor analysis; MOS SF-20 = Medical Outcomes Survey Short Form; NFI = nonnormed fit index; CFI = comparative fit index; AGFI = adjusted goodness-of-fit index; RMSR = root mean square residual; S-B χ^2 = Satorra–Bentler scaled chi-square; CFI* = corrected CFI.

revised scale also to be associated with cardiovascular disease, this link should be empirically verified. To achieve this, a longitudinal design is needed; until then, no definitive conclusions about external validity can be drawn. In statistical terms, it has been determined that the revised Need for Control scale is a one-dimensional construct. We may also conclude that the scale measures the same construct as the original scale, because both have comparable correlations with other variables (see Table 7) as well as a high intercorrelation. However, the revised subscale has better psychometric characteristics and is therefore to be preferred. We strongly recommended that future studies should apply a Mokken analysis to determine the reliability of the dichotomous items, as presented in this article. This analysis method is superior to other more traditional techniques using Cronbach's alpha (see the *Data Analysis* section of the Method section).

Our assessment of congruent validity and content validity allows a few conclusions. The fit of a model in which the Effort–Reward Imbalance subscales load on one common factor is adequate, indicating congruence (see Table 6). The results also show that reward and extrinsic effort are slightly correlated (albeit negatively), suggesting that extrinsic effort and reward may have some commonality (see Table 5). A model (see Table 8), in which the Effort–Reward Imbalance and MOS SF-20 subscales were only allowed second-order covariance, did not fit the data. Therefore, we concluded that covariance between the factors should be allowed, if the model were to fit. This suggests that the scales and subscales of the Effort–Reward Imbalance Questionnaire are interrelated and a closer look at the formulation of the items may partly explain this overlap. The participants were asked to rate "how distressing" certain effortful and rewarding aspects of their work were, and because in both cases a distress rating was given, it seems plausible that effort and reward do have something in common. It could be argued that what is actually measured is the amount of distress caused by effort and reward rather than perceived effort and reward during work. Future research should focus on this conceptual overlap and also try to identify exactly which dimensions of the model are related with health functioning. Efforts should be made to reformulate the items, avoiding the conceptual overlap with distress. Items should be constructed so that they are less susceptible to individual variation. In this respect, Frese and Zapf (1988) proposed formulating items using indexes of frequency rather than reflecting subjective feelings. For example, in the present study,

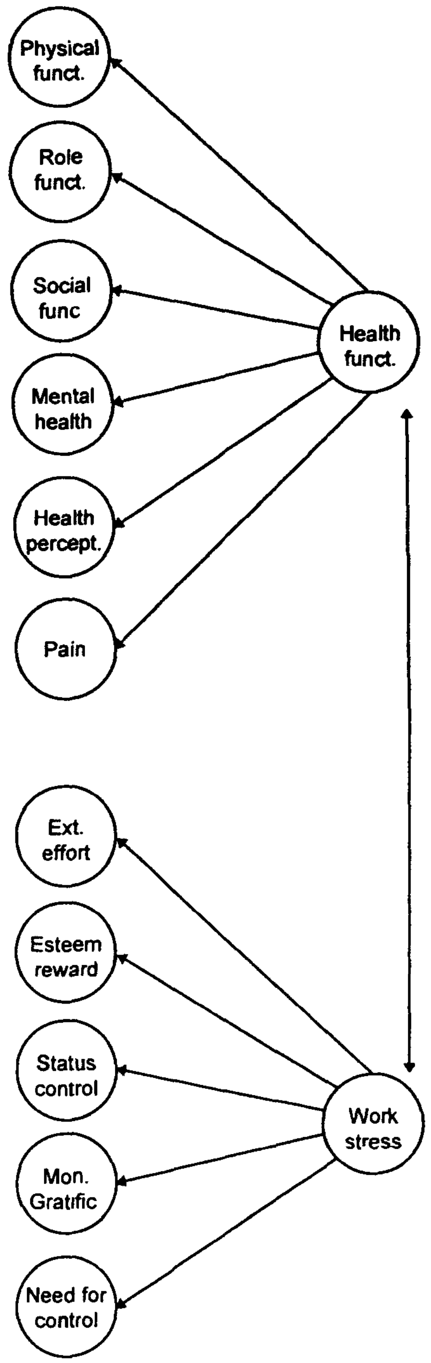

Figure 2 Representation of the model used to determine content validity Covariance is allowed between the second-order factors health functioning and work stress (see M2 in Table 8) Items and error terms are not shown in the figure funct = functioning, percept = perception. Ext = extrinsic. Mon Gratific = monetary gratification

participants had to reply to the effort item "I have constant time pressure due to heavy workload" on a severity scale ranging from *not at all distressed* (1) to *very distressed* (4). Instead of rating their perceived level of distress, the participants could rate how often they experienced time pressure (e.g., everyday, once a week, or once a month).

A drawback of the present study is that the content validity of the Effort–Reward Imbalance instrument is best tested by more objective measures like biomedical events (e.g., myocardial infarct and other vascular incidents), physiological changes (autonomic activation), and performance data (measurement of work output, etc.) and not solely by psychometric analyses as is the case here. Theoretically, a high extrinsic effort and need for control as well as a low reward should be associated with negative emotions and with an activation of the autonomic nervous system. Although this requires intensive measurement techniques, any future attempts to investigate the construct validity should include them. Because the construction of the revised Need for Control scale is based solely on psychometric analyses, determining its relation with external criteria (such as negative emotions and physiological state) might increase the acceptance of the scale. Moreover, this will increase the validity of the scale both nationally and internationally.

Despite these remarks, we may conclude that effort and reward are constructs that measure extrinsic factors in the work environment, such as interruptions and disturbances during work, namely, reward (e.g., salary and job prospects). Socioemotional and motivational aspects that determine need for control are also part of the global effort–reward imbalance construct and measure intrinsic factors. These extrinsic and intrinsic aspects of the work environment have been shown to determine persistence under environmental demands and may also determine an individual's perception of health complaints (Watson & Pennebaker, 1989).

In sum, the following recommendations can be given to increase the instrument's practical use in future studies: (a) Job security should be reflected as a separate dimension in future adaptations of the Effort–Reward Imbalance Questionnaire. (b) The Need for Control scale has been shortened. Its reliability should be assessed by performing a Mokken analysis Its predictive validity for cardiovascular disease should be determined in a longitudinal study. (c) The items of the Effort and Reward subscales should be rephrased to decrease the variability caused by differences in subjective perceptions between participants. Rather than asking subjective rating of work experiences, the items should reflect more quantifiable aspects of the work environment.

Considering the population of the present study, it seems reasonable to conclude that effort–reward imbalance is not solely restricted to blue-collar workers but also to white-collar workers and occupations in the service sector, as has been seen in several international studies using the Effort–Reward Imbalance Questionnaire (Peter & Siegrist, 1997; Siegrist, Bernhart, & Feng, 1990; Siegrist, Peter, Junge, Cremer, & Seidel, 1990). The Dutch translation of the Effort–Reward Imbalance Questionnaire, therefore, seems an adequate self-report measure to assess the level of imbalance due to the work environment. In addition, it may even prove to be relevant for an explanation of self-reported health and well-being. For example, Stansfeld et al. (1998) have shown that aspects of effort and reward at work are related to health functioning. However, their method to determine effort and reward was not standardized, making a replication of their results difficult. We expect an application of the Effort–Reward Imbalance Questionnaire to solve the problems of comparability between studies and other methodological issues (Zapf et al., 1996). Finally, given its predictive power for objective health measures, the Effort–Reward Imbalance Questionnaire seems a promising instrument for future use in this field.

References

Adams, J S (1963). Towards an understanding of inequity. *Advances in Experimental Social Psychology, 67,* 422–436

Adams, J S (1965). Inequity in social exchange *Advances in Experimental and Social Psychology. 2,* 267–299

Bentler, P M (1989) *EQS Structural equations program manual* Los Angeles BMDP Statistical Software

Bentler, P M. & Bonnet, D G. (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin. 88,* 588–606.

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, 88,* 68–74.

Bradburn, N (1969) *The structure of psychological well-being.* Chicago Aldine

Byrne, B M (1994) *Structural equation modelling with EQS and EQS/Windows* Thousand Oaks, CA. Sage.

Caplan, R D , & Jones, K. W (1975) Effects of workload, role ambiguity and Type A personality on anxiety, depression and heart rate *Journal of Applied Psychology, 6,* 713–719

Cooper, C L , & Payne, R (1991) *Causes, coping and consequences of stress at work* Chichester, England Wiley

Debets, P., & Brouwer, E. (1989). *MSP: A program for Mokken scale analysis for polychotomous items.* Groningen, The Netherlands: ProGAMMA.

Dittmann, K., & Matschinger, H. (1982). *Soziale Belastungen, Bewertungsmuster und Stress: Untersuchungen zur Bedeutung psychosozialer Risiken bei der Entstehung von Herzinfarkt. Marburg* [Social strain, esteem reward, and stress· A study of the importance of psychosocial risk factors for the development of heart disease]. Marburg, Germany: Deutches Forschungsverbünde [German Research Foundation].

French, J. R. P., Caplan, R. D., & Harrison, R. V. (1982). *The mechanisms of job stress and strain.* New York: Wiley.

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

Frese, M., & Zapf, D. (1988). Causes, coping and consequences of stress at work. In C. L. Cooper & R. Payne (Eds.), *Methodological issues in the study of work stress: Objective vs. subjective measurement of work stress and the question of longitudinal studies* (pp. 375–412). Chichester, England: Wiley.

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign.* Reading, MA: Addison-Wesley.

Hartley, J., Jacobson, D., Klandersmans, B., & van Vuren, T. (1991). *Job insecurity: Coping with jobs at risk.* London. Sage.

Homans, G. (1961). *Social behavior: Its elementary forms.* New York: Harcourt, Brace, World.

Hu, L., & Bentler, P. M. (1995). Structural equation modelling. In R. H. Hoyle (Ed.), *Evaluating model fit* (pp. 76–99). Thousand Oaks, CA: Sage.

Kahn, R. A (1981). *Work and health.* New York: Wiley.

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24,* 285–311

Karasek, R. A., & Theorell, T. (1990). *Healthy work. Stress, productivity, and the reconstruction of working life.* New York. Basic Books.

Kasl, S V. (1996). The influence of the work environment on cardiovascular health: A historical, conceptual and methodological perspective. *Journal of Occupational Psychology, 1,* 42–56.

Kempen, G I. J. M. (1992a). Het meten van een Nederlandse versie van de MOS-schaal [Measuring a Dutch version of the MOS-scale]. *Tijdschrift Voor Gerontologie en Geriatrie, 23,* 132–140.

Kempen, G. I. J. M. (1992b). The MOS Short-Form General Health Survey· Single item vs multiple measures of health related quality of life. some nuances. *Psychological Reports, 70,* 608–610.

Kempen, G I J. M., Brilman, E I., & Ormel, J (1993). Ouder worden '93 [Growing older '93] Utrecht, The Netherlands· Instituut voor Zorg en Welzijn.

Marmot, M. G. (1994). Work and other factors influencing coronary health and sickness absene. *Work and Stress, 8,* 191–201

Marmot, M. G., & Theorell, T. (1988) Social class and cardiovascular disease· The contribution of work *International Journal of Health Services, 18,* 659–674.

Matschinger, H , Siegrist, J., Siegrist, K., & Ditmann, K. H. (1986). Type A as a coping career: Toward a conceptual and methodological redefinition. In T H. Schmidt, T. M. Dembroski, & G. Blumchen (Eds), *Biological and* *psychological factors in cardiovascular disease* (pp. 104–126). Berlin: Springer-Verlag,

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague, The Netherlands: Mouton.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Physiological Measurement, 6*, 417–430.

Moorer, P., & Suurmeijer, T P B. M. (1993). A study of the unidimensionality and cumulativeness of the MOS Short-Form General Health Survey. *Psychological Reports, 73*. 1324–1326.

Niemoller, B., Schuur, W. H. v., Stokman, F. N. (1980). *Stochastic cumulative scaling Mokken scale, Mokken test (STAP user's manual)*. Amsterdam: University of Amsterdam Press.

Nunnaly, N. C. (1978). *Psychometric theory*. New York: McGraw-Hill

Parkes, K. R (1994). Personality and coping as moderators of work stress processes: Models, methods and measures [Special issue: A healthier work environment]. *Work and Stress, 8*, 110–129.

Peter, R., & Siegrist, J (1997). Chronic work stress, sickness absence, and hypertension in middle managers: General or specific sociological explanations. *Social Science and Medicine, 45*, 1111–1120

Rasch, G (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark. Danish Institute for Educational Research.

Satorra, A., & Bentler, P. M. (1988). *Scaling corrections for chi square statistics in covariance structure analysis*. Alexandria, VA· American Statistical Association.

Sauter, S. C., Hurrell, J. J, & Cooper, C. L. (1989). *Job control and worker health*. Chichester, England· Wiley.

Schnall, P. L, Landsbergis, P. A, & Becker, D. (1994). Job strain and cardiovascular disease. *Annual Review of Public Health, 15*, 381–411

Siegrist, J (1996a) Adverse health effects of high effort–low reward conditions. *Journal of Occupational Health Psychology, 1*, 27–41

Siegrist, J. (1996b) *Soziale Krisen und Gesundheit* [Social crises and health]. Gottingen, Germany Hogrefe

Siegrist, J., Bernhardt, R, & Feng, Z. (1990). Socioeconomic differences in cardiovascular risk factors in China *International Journal of Epidemiology, 19*, 905–910

Siegrist, J., & Peter, R. (1994). Job stressors and coping characteristics in work-related disease: Issues of validity. *Work and Stress, 8*, 130–140.

Siegrist, J., Peter, R., Junge, A., Cremer, P., & Seidel, D. (1990). Low status control, high effort at work and ischemic heart disease: Prospective evidence from blue-collar men. *Social Science and Medicine, 31*, 1127–1134.

Sijtsma, K., Debets, P., & Molenaar, I. W. (1990). Mokken scale analysis for polychotomous items: Theory, a computer program, and an empirical application. *Quality and Quantity, 24*, 173–188.

Stansfeld, 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, 60*, 247–255.

Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form general health survey. *Medical Care, 26*, 724–735.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*. 1–10.

Veit, C T., & Ware, J. E. (1983). The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology, 51*, 730–772.

Walster, E., Berscheid, E., & Walster, G. W. (1973). New directions in equity research. *Journal of Personality and Social Psychology, 25*. 151–176.

Warr, P. B. (1987). *Work unemployment and mental health*. Oxford, England. Oxford University Press.

Warr, P. B. (1994). A conceptual framework for the study of work and mental health. *Work and Stress, 8*, 84–97.

Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity *Psychological Review, 96*, 234–254.

Zapf, D, Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research· A review of the literature with reference to methodological issues. *Journal of Occupational Health, 1*, 145–169.

Received March 30, 1998
Revision received December 18, 1998
Accepted March 19, 1999 ■