Chapter 2
The Measurement of Effort-Reward Imbalance (ERI) at Work

Diego Montano, Jian Li, and Johannes Siegrist

2.1 Introduction

The measurement process in the social and behavioral sciences is more problematic than in the natural and basic sciences. Major measurement problems associated with social and behavioral phenomena are related to the high complexity of biological systems, the difficulty of defining unequivocal objects of measurement, and the practical cost and time constraints concerning data collection from different sources. Moreover, in many cases, calibrated measurement standards that are required for unequivocal quantification, comparison and replication of data are not available. Despite these limitations remarkable methodological advances were achieved in these sciences, including the assessment of people's personal experiences, attitudes, and behaviors (King et al. 2004).

One such advance concerns the modeling of constructs by means of *latent variables* and their measurement by a set of standardized indicators. There are two main approaches towards assessing these indicators, observational and self-report data collection. Either approach has its strengths and weaknesses. To measure the construct 'effort-reward imbalance' (ERI) in terms of a *systematic observation* of

D. Montano (✉)
Institute for Medical Psychology and Sociology, University of Goettingen, Goettingen, Germany
e-mail: diego.montano@med.uni-goettingen.de

J. Li
Institute of Occupational and Social Medicine, University of Düsseldorf, Düsseldorf, Germany
e-mail: jian.li@uni-duesseldorf.de

J. Siegrist
Faculty of Medicine, University of Düsseldorf, Düsseldorf, Germany
e-mail: Johannes.siegrist@med.uni-duesseldorf.de

J. Siegrist, M. Wahrendorf (eds.), *Work Stress and Health in a Globalized Economy*, Aligning Perspectives on Health, Safety and Well-Being,
DOI 10.1007/978-3-319-32937-6_2

the efforts and rewards people experience at work has the advantage of providing information that may be unbiased by subjective interpretations. Yet, distinct components of the model (e.g. esteem reward, over-commitment) are hardly observable, and due to the narrow time window of observation relevant occurrences may be bypassed or interpretation biases of external observers may limit the validity of assessment. Additional problems concern the high costs of application and training, and the limited size of samples to which this technique can be applied (Karasek and Theorell 1990; Rau et al. 2010). *Self-report data* are more easily collected, and there are no limitations of the sample size. However, the validity of this measurement has been repeatedly questioned given several sources of systematic bias, such as the risk of distortion of information, the impact of distinct personality traits and reporting styles, and uncontrolled contextual influences on the assessment process (Kahneman et al. 2004; Sudman et al. 1996). On balance, self-report data have the advantage to reflect personal experience, a crucial prerequisite of human stress research, and they cover long periods of the person's living and working conditions.

Different methods of collecting self-report data have been applied to the measurement of effort-reward imbalance at work. They include non-standardized qualitative interviews (McGillis Hall and Kiesners 2005), event momentary assessment (Johnston et al. 2013; Johnston 2006), standardized computer-assisted personal or telephone interviews that may or may not include psychometrically validated scales (Wege et al. 2008), and questionnaires containing psychometrically validated scales answered either under controlled conditions or via mail or online (Fekete et al. 2014).

The *development and test of ERI indicators* has taken a number of years and underwent several changes. In a first phase, data were collected from healthy workers as well as from distinct groups of people with chronic disease (in particular coronary heart disease), by using semi-structured interviews with descriptive and evaluative questions measuring 'effort' and 'reward', and additionally by applying a pool of dichotomous items assessing 'over-commitment'. It is important to note that descriptive information was validated by contextual data where available (for a detailed description see Siegrist (1996b), for a summary description see the main ERI reference paper Siegrist (1996a)).

In a second phase, based on the results of these early studies, each component of the construct was assessed by a series of *Likert-scaled items* reflecting the questions with highest explanatory power. Items and related scales underwent psychometric analysis according to classical psychological test statistics, including the assessment of internal consistency of scales, of item-scale correlations, and of scale intercorrelations. The *three scales* were finally composed by six items measuring '*effort*' (five items if 'physical workload' was excluded, e.g. in white-collar surveys), 11 items measuring '*reward*' (containing the three sub-scales 'money and career', 'esteem', 'job security'), and six items measuring the personal coping pattern of '*over-commitment*' (Siegrist et al. 2004). Confirmatory factor analysis was performed by comparing alternative configurations of the model structure, where the formal structure reflecting most closely the theoretical assumptions demonstrated the best fit with data (Rödel et al. 2004). This latter finding was replicated in several studies testing the original and the newly developed short version of the

questionnaire (see text and Fig. 2.1 below). Later on, the original version of the ERI questionnaire was subject to a change of the answering format of items, and an additional, psychometrically validated short version was developed.

These changes of the ERI questionnaire are described in more detail in the next section. Following this, the main ways of testing the research hypotheses of this

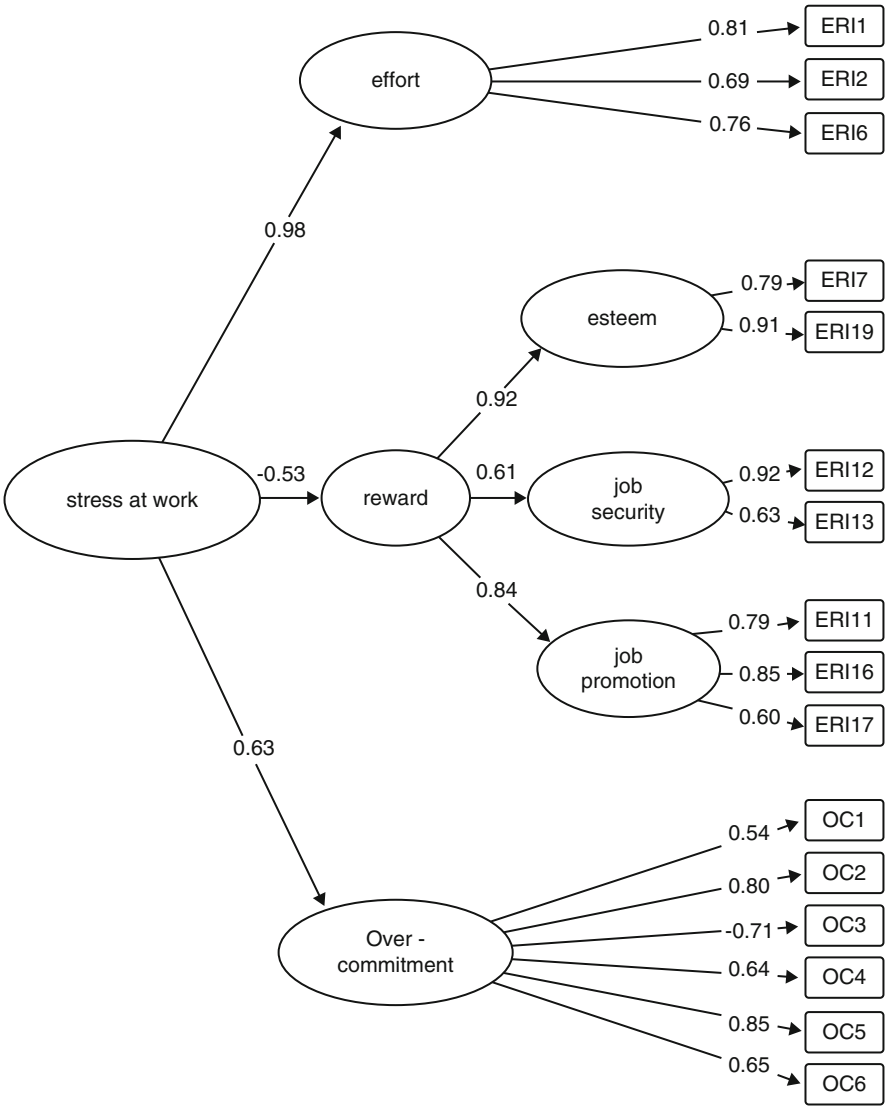

Fig. 2.1 Confirmatory factor analysis testing the ERI-model (short version, source: Leineweber et al. 2010)

model are discussed. In the third section, major critical methodological problems are addressed, and in the final section some future directions of research based on this measurement approach are proposed.

2.2 ERI Scales and Test of the Model's Hypotheses

2.2.1 Towards a One-Step Rating Procedure

When the *original 23 item version* of the ERI questionnaire was developed the prevailing psychological stress theory suggested that cognitive appraisal exerts a decisive impact on the triggering of stress responses (Lazarus and Folkman 1984). According to this assumption a two-step procedure of answering the items was proposed to map the distinction between occurrence of a situation and appraisal of the degree of distress assigned to it. Therefore, first, subjects agree or disagree whether or not the item content describes a typical experience of their work situation. Subsequently, subjects who agree are asked to evaluate to what extent they usually feel distressed by this typical experience (not distressed, somewhat distressed, distressed, very distressed) (Siegrist et al. 2004). As discussed below, this two-step 5-point Likert scale rating procedure turned out to be problematic in psychometric terms, and it produced some difficulties during data collection, especially among low educated people. Therefore, in a further improvement of the ERI questionnaire, a widely used one-step rating procedure was proposed in terms of a 4-point Likert scale where respondents had to indicate to what extent they agree or disagree with the item content (strongly disagree, disagree, agree, strongly agree) (Siegrist et al. 2009). Here, the distinction between occurrence and appraisal was no longer made, and the rating of the degree of stressfulness was replaced by assessing to what extent the item content matches the respondents' typical experience at work. As the item content describes a situation generally appraised as disadvantageous (or advantageous if reverse coding is required) the respondents' answers to the items of each scale are expected to mirror the amount of their *perceived* effort and reward at work.

The new rating procedure was introduced in the context of developing a psychometrically validated *short version of the original questionnaire* (see below). It was also widely used in the short version incorporated in the Survey of Health, Ageing and Retirement in Europe (SHARE) (Siegrist et al. 2007; Wahrendorf and Siegrist 2014). Based on the experience of appropriate psychometric properties of the scales and on the observation that associations with health outcomes were generally well comparable to those derived from the original questionnaire (Fekete et al. 2014; Juvani et al. 2014; Li et al. 2012a, b; Tsutsumi et al. 2008), it was more recently proposed to use this new rating procedure in future applications of the original ERI questionnaire as well (Siegrist et al. 2014). However, one has to take into account that the sum scores of the scales with 4-point answers and those with 5-point answers have to be transformed in order to be comparable.

2.2.2 The Short Version of the Questionnaire

Given increasing demands from research consortia of large-scale cohort studies to provide a short, economic measure of the model, an abbreviated version of the two scales 'effort' and 'reward' was developed in Germany, based on data from the Socioeconomic Panel (Siegrist et al. 2009), and in Sweden, based on data from the Swedish Longitudinal Occupational Survey of Health (Leineweber et al. 2010). The two versions were almost identical, containing ten items (three 'effort' and seven 'reward' items). In either case, a second order model (where reward was represented by the three theoretically defined components; see Fig. 2.1) reached the relatively best level of fit, as indicated by several fit indices. Moreover, the criterion validity was tested by analyzing associations with measures of self-rated health and depressive symptoms. High effort in combination with low reward was strongly and consistently related to poor self-rated health in the cross-sectional study in Germany (Siegrist et al. 2009) and in the longitudinal study in Sweden, where similar associations were additionally observed for depressive symptoms (Leineweber et al. 2010). Importantly, an *identical factorial structure* resulted from the Chinese short version of this questionnaire, and significant associations of effort-reward imbalance with poor physical and mental health were reported in this large community survey in the city of Kunming (Li et al. 2012b). The factorial structure was again confirmed in the short version of the questionnaire applied in a different cohort of German workers (Li et al. 2012a).

In conclusion, based on this evidence, the short ERI questionnaire can be recommended for application in large-scale studies where economic measures are required. However, the original questionnaire covers the model's components in a more comprehensive way and is therefore considered the first choice, in combination with the one-step rating procedure (Siegrist et al. 2014). This conclusion is substantiated by further test-statistical information demonstrating the replication of its factorial structure in different countries (see Table 2.2) and the factorial invariance and stability over time (de Jonge et al. 2008; Rantanen et al. 2012; Törnroos et al. 2014). This latter finding is of course essential for obtaining unbiased estimates of change in indicators of work-related stress (e.g. in intervention studies). Thus, it can be assumed that changes in the parameters of the ERI scales reflect true changes in work characteristics (Törnroos et al. 2014).

2.2.3 Test of the Model's Hypotheses

As described in the previous chapter, the main aim of the ERI model consists in producing new knowledge on associations of stressful working conditions with workers' health outcomes. Considering the model's components three hypotheses were stated. They are visualized in a causal path model in Fig. 2.2 (Siegrist 2002):

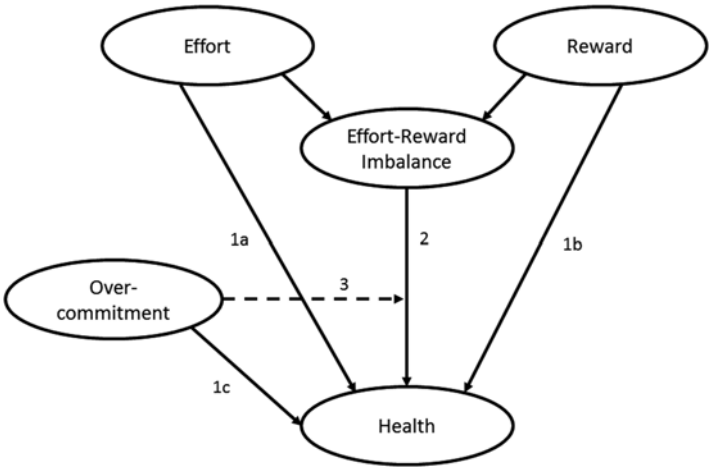

Fig. 2.2 Path diagram representing the causal structure implied by the effort reward imbalance model

1. Each component ('effort', 'reward', 'over-commitment') exerts separate effects on the health outcome under study (paths 1a, 1b, 1c in Fig. 2.2).
2. The size of effect on health produced by a combined measure of high effort and low reward exceeds the size of effect on health produced by each single scale (e.g. as demonstrated by the individually assessed 'effort/reward ratio') (path 2 in Fig. 2.2).
3. The personal coping pattern 'over-commitment' moderates the effect size of effort-reward imbalance on health (interaction term) (path 3 in Fig. 2.2).

It should be stated that two different approaches were proposed towards testing these hypotheses. In the approach established most pervasively in psychology and sociology, estimating the causal structure of latent (or observed) variables, as depicted in Fig. 2.2, requires the identification of main and interaction effects of effort, reward, and over-commitment. Thus, testing a multiplicative term of the scales has been proposed as a way of analyzing the model's hypotheses (van Vegchel et al. 2005).

In epidemiological research the size of effect on health produced by a predictor is estimated in terms of relative risk, hazard ratio, or odds ratio using multivariate regression analysis. In this regard, it is of interest to compare the effect size of the model's single components with the effect size of a *combined measure of effort and reward*, based on an algorithm that *quantifies their imbalance at individual level*. This algorithm is of theoretical interest as well. According to the theory of affective information processing (Ledoux 1989) a potentially unfavorable trade-off between the costs and gains experienced in everyday working life is rarely subject to conscious computational processing. Rather, negative emotions aroused by the recur-

rent experience of non-reciprocal exchange at work may bypass conscious awareness, as is the case for a substantial part of affective processing in general (Siegrist et al. 2004). An investigator-based algorithm quantifying this mismatch may capture part of respective strain reactions that would be missed if measurement of imbalance were based on the working person's explicit trade-off of cost and gain at work (Siegrist et al. 2004). In fact, a substantial number of epidemiological studies provide empirical support of this theoretical assumption (see several chapters in this book).

2.3 Critical Aspects Concerning the Psychometric Properties of the ERI Scales

Despite extensive psychometric analyses performed with the ERI scales in international research and despite a robust body of empirical evidence that the model's components contribute to the prediction of different stress-related disorders, measures of functioning, and well-being, there remain some critical methodological aspects that need to be discussed. In this section, three such critical aspects are given special attention. First, as mentioned above, the shift from a two-stage response pattern of the items defining the scales to a one-step rating procedure has a serious impact on the distribution of answers and, subsequently, on comparing the results between the two procedures. Second, problems of defining a cut-point or critical threshold in the distribution of sum scores of the scales are discussed. Finally, some challenges of trans-cultural application of the scales are addressed.

2.3.1 Comparability of Different Rating Procedures

Even though the changes mentioned above can be regarded as improvements, they may at the same time compromise the comparability of results across studies insofar as they affect fundamental statistical parameters such as means, variances and covariances. In other words, a direct comparison of distribution parameters and effect sizes across studies requires taking into account the type of rating procedure applied in the studies. In Table 2.1 the two rating procedures are displayed. It should be noted that some studies applying the 1-step rating procedure offer a 5-point Likert scale instead of a 4-point scale (Hintsanen et al. 2007; Kivimäki et al. 2007; Li et al. 2012a; Yokoyama et al. 2014).

Nonetheless, as pointed out by Kurioka et al. (2013), changing the rating procedure may shift the distribution of the effort, reward and *effort-reward ratio* (ER ratio) scores. As a result, applying the two-step or the one-step procedure is likely to have two major side effects: The first one concerns statistical properties of the scales. From a mathematical point of view, the main reason accounting for the

Table 2.1 Common rating procedure formats of the ERI questionnaire

	2-step rating procedure with five categories (original procedure). Respondents are asked whether the item describes a situation encountered at work (does not apply/apply). If the item applies, respondents are then asked to rate on a 4-point Likert scale the extent to which the work situation is experienced as distressful		**1-step rating procedure with four categories (revised procedure).** Respondents are asked to rate on a 4-point Likert scale the extent to which they agree or disagree with the item content	
Construct assessed	Presence or absence of stressful work environment component reflected in item content, and subjective experience of emotional distress		Subjectively perceived stressfulness of item content	
Categories	**Item response formats**	**Score**	**Item response format**	**Score**
Category 1	Disagree/agree	1	Strongly disagree	1
Category 2	Agree, but I am not at all distressed	2	Disagree	2
Category 3	Agree, and I am somewhat distressed	3	Agree	3
Category 4	Agree, and I am distressed	4	Strongly agree	4
Category 5	Agree, and I am very distressed	5		

distribution shifts in all ERI scales is the fact that the frequencies of categories 1 and 2 for all items in the two-step procedure are higher in comparison to the other categories. In contrast, the frequencies of categories 2 and 3 in the one-step procedure are much higher than those of categories 1 and 4. As a consequence, the ratio of effort and reward scores tends to be much lower in the two-step than in the one-step procedure. The second effect relates to the object of measurement. From a psychometric point of view, changing the rating procedure from a two-step to a one-step format implies focusing exclusively on the working persons' perception of the degree of fit between a potentially stressful aspect of the work environment, as illustrated by the item content, and their respective personal experience (see above Sect. 2.2.1).

In order to illustrate how the *different rating procedures* may affect the distribution of ERI scales, we simulated a dummy dataset containing 1000 observations by randomly sampling from the items range (i.e. 1–5 and 1–4, respectively), and evaluated the impact of the rating procedures on the distribution of scales, based on the original version of the scales. We took into account the unequal selection probability of categories in each procedure described by Kurioka et al. (2013), i.e. a larger probability of choosing categories 1 and 2 in the two-step procedure, and a larger probability of choosing categories 2 and 3 in the one-step procedure (see Table 2.1 for details). We simulated the responses to 16 items on a 5-point Likert scale. As depicted in Fig. 2.3, the distribution of the ERI scores in the two-step procedure

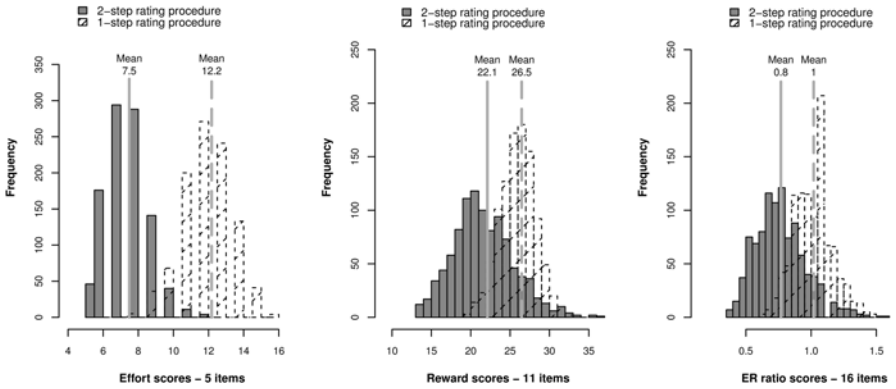

Fig. 2.3 Comparison of the distribution of the ERI scales according to the one- and two-step rating procedures with a simulated dataset containing 1000 observations. Sampling probabilities of categories 1 and 2 for the 2-step procedure, and of categories 2 and 3 for the 1-step procedure are higher (0.9 and 0.7; 0.7 and 0.6, respectively). The *vertical lines* on each histogram indicate the sample mean of the distribution under the two-step procedure (*continuous line*) and the one-step procedure (*dotted line*)

is consistently shifted left to the distribution of scores in the one-step procedure. Notice that this phenomenon was reported by Kurioka et al. (2013) with survey data.

From a statistical perspective, however, the shifting of scales suggests that the one-step procedure may be interpreted as a linear transformation of the scales obtained from the two-step procedure. Since linear transformations of the form $X = aX + b$ do not affect the correlation between two random variables, say between variables X and Y, the inferences obtained from the ERI scales by means of the one- or two-step rating procedure should be to some extent comparable for both long and short versions in spite of the distribution shifts observed in empirical research.

2.3.2 *Critical Thresholds of Scales?*

The ERI scales were conceptualized as a method of assessing stressful work in employed populations and of estimating associated health risks. Therefore, the higher the *sum score of the scales*, including the ratio, within a distinct population group, the higher the probability of occurrence of a stress-related health outcome. The scales were not constructed as a diagnostic tool to assess individual risk. Thus, clinically validated cut-points in terms of classical psychological test theory are not provided. It is nevertheless of interest to explore whether the hypothesis of a *linear relationship* always holds true or whether effect sizes on health vary according to a *distinct threshold*. For instance, sensitivity and specificity of cut-points have been estimated comparing the scores of patient groups and healthy control groups (Lehr et al. 2010). In fact, in earlier publications the authors of the ERI questionnaire

proposed to use a cut-point of the effort-reward (ER) ratio, where scores beyond 1.0 are thought to represent a critical condition of high cost/low gain whereas scores below this threshold are assumed to indicate absence of risk (Siegrist et al. 2004). Although this idea is intriguing it turns out to be problematic in terms of measurement theory. This argument is discussed in more depth in the following paragraph. Here, it is important to stress that it is generally recommended to *analyze continuous data of the scales* which then might be re-classified, e.g. as quartiles, for inclusion in logistic regression analysis.

The *assumption of linearity* deserves a further theoretical comment. Distinct from this model, equity theory predicts a curvilinear relationship between equity and strain where high reward in combination with low effort raises negative emotions as well (Adams 1965). However, in keeping with the propositions of prospect theory (Kahneman and Tversky 1979) and of the conservation of resources theory (Hobfoll 1989), the ERI model posits that threat to, or experience of loss following high effort spent matters most for health, due to the intensity of reactions following the violation of one of the most profound principles of interpersonal exchange (Gouldner 1960). Recent evidence from neuroscience research supports this argument (Hernandez Lallement et al. 2013; Tricomi et al. 2010). Future analyses might nevertheless explore a J-shaped relationship between effort-reward imbalance and health, where obviously exaggerated, unjustified rewards in combination with low effort are expected to stimulate unpleasant feelings of social disapproval.

Turning to the issue of *defining a threshold*, specifically based on scores obtained from the ER ratio, *several risks* have to be taken into account. First, assuming that the two units of the effort and reward scales are strictly comparable, an algorithm dividing 'effort' by 'reward' by adjusting for unequal item numbers results in a different prevalence of persons at risk, depending on whether the two-step or the one-step answer format has been applied (see Fig. 2.3 above). The simulated scores suggest a high number of 'false positive' classifications in the one-step procedure. In other words, additional ROC curve analysis needs to be performed in order to adjust the cut-off point accordingly. This has been done e.g. by Kurioka et al. (2013), whereas in the Msaouel et al. (2012) study this was not accomplished, with the consequence of reporting an excessively high prevalence of Greek health care workers exposed to stressful work (80.7 %).

Second, a problem of underestimation of ER ratio with health outcomes may arise in case of a high prevalence of the exposure. For instance, in a study of hotel room cleaners, Krause et al. (2010) reported a mean ER ratio of 1.3 (SD 0.9), but conducted analyses based on a cut-point 1.0. This procedure resulted in weak effect sizes in terms of odds ratios, compared to additional analyses based on quartiles of the ER ratio where odds ratios were substantially higher (see Yokoyama et al. (2014) for a similar underestimation). On the other hand, in cases where the sample distribution of the ER ratio is positively skewed (i.e. lower values are much more frequent), the test specificity largely increases if the ER ratio is dichotomized at 1.0. As a consequence, not only the strength of the association between the effort-reward imbalance scale and health will be overestimated, but also the corresponding standard errors of estimates (see Lehr et al. (2010) for a detailed cut-point analysis).

A third criticism relates to the substantial loss of information by reducing the data set to a dichotomized variable. Therefore, as mentioned above, analyses based on thresholds should be performed very carefully, considering the specific distribution of scales and eventually performing additional ROC analysis. In all cases, continuous data analysis should be the first strategy (Niedhammer et al. 2004; Royston et al. 2006), and evidence so far suggests that scores in the upper quartile of the distribution often indicate susceptibility to elevated health risks.

2.3.3 Trans-cultural Application of Scales

To translate and apply a survey instrument in a different language and socio-cultural context is considered a highly challenging task. First, the *translation process* itself has to meet defined quality criteria including forward- and backward translation by two independent native speakers with bi-lingual competence, ascertaining the provision of language-specific equivalence of meaning (Harkness et al. 2010). Today, different tools of language management utility are available, but the fine-tuning of translation usually requires elaborated personal experience with respective language. Once this translation process has been achieved, discrepancies in meaning and uncertainties are discussed among experts. In case of the ERI questionnaire they often involved consultations with the author of the instrument. As a next step, the translated questionnaire needs pre-testing and related data analysis, resulting eventually in further revision before being ready for routine assessment. With the development of the *Differential Item Functioning* (DIF) approach (Holland and Wainer 1993) it has become possible to estimate in cross-cultural comparative investigations to what extent insufficient item functioning is attributable to difficulties in translation – as an example Choi et al. (2009).

A related second, much more challenging problem of applying standardized questionnaires across countries concerns the *cultural variation of* the *meaning* of specific terms, verbal expressions or whole sentences, and the divergences of underlying attitudes and social norms. This fundamental problem has been discussed in major publications (Burke 2010; Hofstede 2001; House et al. 2004), and cannot be analyzed in detail here. Yet, several examples of problems in ascertaining an equitable meaning of translated items of the ERI scales may illustrate the case.

In a recently developed Arabic version of the ERI questionnaire, the item content of "job security" reflects two meanings, "employment security" and "employees' physical protection at work". While the latter notion is not at all intended in the original version, it is more widely rooted in the experiences of Arabic people and their culture (Almadi et al. 2013). A second example refers to items of the 'effort' scale. When testing item functioning between a Japanese and a Dutch sample, the item "I am often pressured to work overtime" strongly deviated from the expected convergence (Tsutsumi et al. 2009). This observation may point to the specific cultural background of excessive work obligations in Japan, as reflected by the Karoshi phenomenon, among others (Yang et al. 2015). Regarding the results of factorial

analysis in several investigations conducted in Asian countries, the item "I get easily overwhelmed by time pressures at work" displayed loadings of similar strength on the factor 'over-commitment' and on the factor 'effort' (Almadi et al. 2013; Eum et al. 2007; Li et al. 2005). Thus, a clear distinction between extrinsic and intrinsic aspects of demanding situations seems to be prevented by respective cultural backgrounds.

These challenges have to be taken into account, in particular in a measurement approach that is expected to contribute to the explanation of work-related health risks in a globalized economy, as is the case with ERI. As indicated in several chapters of this book (specifically part III) it is nevertheless rather remarkable to see the degree of consistency of associations of stressful work in terms of ERI, as measured in different languages, and poor health across different countries not only within a single continent (e.g. Europe), but also between different continents (e.g. Siegrist et al. (2012)).

So far, to our knowledge, the *ERI questionnaire* has been *translated and psychometrically validated in 14 different languages*. The main psychometric properties are given in Table 2.2. In general, the internal consistency of the scales is satisfactory, and the factorial structure of the scales has been replicated in the majority of samples by confirmatory factor analysis. Moreover, criterion validity has been explored in several studies, confirming the utility of the scales in explaining health risks.

Additional translations of the ERI questionnaire were performed in several languages, but respective publications did not provide detailed data on psychometric validation, particularly on confirmatory factor analysis. These language versions are Arabic (Almadi et al. 2013), Czech, Hungarian, Lithuanian, Polish (Pikhart et al. 2001), Greek (Msaouel et al. 2012), Italian (Zurlo et al. 2010), Japanese (Tsutsumi et al. 2001), Mongolian (Bagaajav et al. 2011), Russian (Pikhart et al. 2004), Sinhala (Gamage and Seneviratne 2015), Slovakian (Hasselhorn et al. 2004), and Spanish (Macias Robles et al. 2003).

Notably, as the first version out of Europe, the Japanese version of the ERI questionnaire deserves special attention. In addition to the satisfactory psychometric properties (Tsutsumi et al. 2001), the Japanese version also demonstrated valid responsiveness to organizational change over time (Tsutsumi et al. 2002). Methodologically, the Japanese team was the first to test the two different rating procedures in the standard version (Fekete et al. 2014; Juvani et al. 2014; Li et al. 2012a, b; Tsutsumi et al. 2008) and the short version (Kurioka et al. (2013) (for more details see Sect. 2.3.1; see also Chap. 8). Moreover, the cross-cultural comparability of the ERI questionnaire was critically evaluated by means of differential item functioning (DIF) analyses (Tsutsumi et al. 2009).

In conclusion, despite far-reaching challenges of conducting trans-cultural comparative research the findings based on the ERI questionnaire provide convincing evidence that these comparisons offer reliable and valid information. Still, further testing and critical appraisal of available evidence will be required.

Table 2.2 Overview of psychometric properties of the ERI questionnaire in different language versions (in alphabetic order)

Language version	Population	Cronbach's α coefficients	Confirmatory factor analyses	Discriminant validity
Chinese (Li et al. 2005)	Healthcare workers, 192 men, 608 women	Effort: 0.78 Reward: 0.81 Over-commitment: 0.74	Goodness-of-fit index > =0.90	Gender, age, education
Danish (Weyers et al. 2006)	367 female nurses and nurses' aides	Effort: 0.71 Reward: 0.78 Over-commitment: 0.76	Goodness-of-fit index > =0.90[a]	–
Dutch (Hanson et al. 2000)	775 employees from four companies, 82 % men, 18 % women	Effort: 0.71 Reward: 0.70–0.77 Over-commitment: 0.82	Goodness-of-fit index > =0.90[a]	–
English-British (Siegrist et al. 2004)	Whitehall II Study (civil servants) 2783 men, 914 women	Effort: 0.73 men, 0.76 women Reward: 0.83 men, 0.84 women Over-commitment: 0.81 men, 0.82 women	Goodness-of-fit index > =0.90[a]	Gender, age, education
Farsi (Yadegarfar et al. 2013)	227 male employees of Iran Polyacryl Corporation	Effort: 0.70 Reward: 0.88 Over-commitment: 0.72	Goodness-of-fit index > =0.90[a]	–
Finnish (Kinnunen et al. 2007)	1301 managers, 70 % men, 30 % women	Effort: 0.83 Reward: 0.87 Over-commitment: 0.76	Comparative fit index > =0.90[a]	Gender, age, marital status, managerial levels
French (Siegrist et al. 2004)	GAZEL (French National Electric and Gas Company) 7251 men, 2923 women	Effort: 0.75 men, 0.75 women Reward: 0.86 men, 0.88 women Over-commitment: 0.79 men, 0.79 women	Goodness-of-fit index > =0.90[a]	Gender, age, education

(continued)

Table 2.2 (continued)

Language version	Population	Cronbach's α coefficients	Confirmatory factor analyses	Discriminant validity
French-Belgian (Siegrist et al. 2004)	Somstress study (4 companies across Belgium)	Effort: 0.64 men, 0.72 women	Goodness-of-fit index > =0.90[a]	Gender, age, education
	2055 men, 1,739 women	Reward: 0.78 men, 0.77 women		
		Over-commitment: 0.81 men, 0.80 women		
German (Siegrist et al. 2004)	Public transport employees	Effort: 0.68 men, 0.61 women	Goodness-of-fit index > =0.90[a]	Gender, age, education
	267 men, 48 women	Reward: 0.86 men, 0.87 women		
		Over-commitment: 0.74 men, 0.64 women		
Korean (Eum et al. 2007)	908 male petrochemical workers	Effort: 0.71	Goodness-of-fit index > =0.90	Age, education, employment grade
		Reward: 0.86		
		Over-commitment: 0.75		
Norwegian (Lau 2008)	1803 municipality employees, 368 men, 1433 women	Effort: 0.72	Goodness-of-fit index > =0.90[a]	Gender, age, education, occupational groups
		Reward: 0.78		
		Over-commitment: 0.76		
Portuguese-Brazilian (Griep et al. 2009)	1509 nursing personnel, 86.5 % women	Effort: 0.73	Goodness-of-fit index > =0.90[a]	–
		Reward: 0.76		
		Over-commitment: 0.75		
Spanish in six Latin-American countries (Argentina, Chile, Colombia, Mexico, Peru, and Venezuela) (Juarez-Garcia et al. 2015)	1292 health professionals, 283 men, 1009 women	Effort: 0.80	Goodness-of-fit index > =0.90[a]	Age, marital status, education
		Reward: 0.86		
		Over-commitment: 0.73		

(continued)

Table 2.2 (continued)

Language version	Population	Cronbach's α coefficients	Confirmatory factor analyses	Discriminant validity
Swedish (Siegrist et al. 2004)	WOLF (several companies representing different sectors in the Northern region of Sweden, Norrland)	Effort: 0.71 men, 0.78 women	Goodness-of-fit index > =0.90[a]	Gender, age, education
	738 men, 222 women	Reward: 0.79 men, 0.70 women		
		Over-commitment: 0.80 men, 0.80 women		
Thai (Buapetch et al. 2008)	828 garment factory workers, 137 men, 691 women	Effort: 0.77	Goodness-of-fit index > =0.90	–
		Reward: 0.81		
		Over-commitment: 0.66		

[a]Indicates the second-order factor analysis on reward scale

2.4 Some Suggestions for Future Developments

Up to now, the ERI questionnaire has been applied in its original or shortened version in a large number of epidemiological investigations and in many experimental or naturalistic studies in several parts of the world. *Uniform application of a measurement* approach must be considered a relevant *scientific goal*. In earlier years several important publications testing the ERI model were based on proxy measures, compromising the comparability of findings. It is therefore crucial to *apply psychometrically validated*, *standardized ERI measures* in future research. Despite this call for increased comparability there must also be room for further improvements and innovations of a method that was originally developed more than two decades ago. In this section, some respective suggestions are given.

One suggestion concerns the *specification of the measurement model*. Two major models have been largely debated in the psychometric literature: the reflexive and the formative measurement models. Whereas reflexive models assume that the latent variable causes the observed changes in the items or indicators (classical test theory), formative models postulate, in contrast, that indicators cause changes on the latent variable scale (Bollen and Lennox 1991; Diamantopoulos et al. 2008). Furthermore, indicators in formative models are not required to correlate, and may be rather interpreted in the framework of conventional multiple regression models

as independent variables determining the values of the latent variable. Hence, the usual practice of summing up the item scores for calculating an unweighted overall score may be inadequate in formative scales, since some items may have a much larger influence in the overall levels of the latent variable (Fayers et al. 1997).

Following classical test theory, the latent variables of the ERI questionnaire were originally developed within the framework of reflexive models (Hanson et al. 2000; Peter and Siegrist 1997; Siegrist 1996b). This approach has recently been challenged by Pitts (2014) who pointed out that the latent variables 'effort' and 'reward' do not necessarily imply changes in the corresponding items, and should therefore be analyzed by means of a formative measurement model in order to avoid potential misspecification of the measurement model. However, the assumption of a formative model seems to suffer from important epistemological and methodological problems including the difficulties of assessing construct validity and model identification, measurement error, and unchallenged causality assumptions (Edwards 2011). Thus, given the fact that the ERI scales have shown acceptable levels of reliability, criterion and discriminant validity, satisfactory predictive power in occupational health and social epidemiology research, and factorial group and time invariance (de Jonge et al. 2008; Rantanen et al. 2012; Törnroos et al. 2014), it seems likely that their summary statistics based on the reflexive measurement model provide robust scientific information, especially so if the critical aspects discussed above are taken into account.

Critical aspects may be extended to include the specification of a particular psychometric model, e.g. item-response theory, latent class variables, formative or reflexive models (Borsboom 2006; Borsboom et al. 2004; Hayduk and Littvay 2012). Moreover, in line with modern psychometric test theory, criterion-based item scoring can be applied to compute scale scores that are independent of empirical score distributions and, thus, well comparable across studies. One such approach was successfully applied to the ERI questionnaire (Hadzibajramovic et al. 2015).

Even most sophisticated methodological approaches cannot resolve the problem that *self-reported data* reflect *perceptions of reality* rather than an 'objective' reality. Yet, for several reasons this problem should not be overstated (Kompier 2005). First, major traditions in social and behavioral sciences, including human stress research, argue that *subject's definitions* of what is real *matter most* because these definitions "are real in their consequences" (Thomas and Thomas (1928), 571f.). Second, a substantial body of evidence indicates that much of what is reported as experienced reality through interviews or questionnaires is in line with observable or otherwise independently confirmed facts. This is ascertained by *triangulation*, e.g. by comparing reported information with biographical data, administrative records, validation by significant others etc. (Siegrist et al. 1988). Another way of tackling the problem of potential distortion inherent in subjective reports is the *aggregation of individual data* at the level of homogenous groups and to compare results obtained from aggregate measures as predictors with the results obtained from individual measures as predictors. For instance, in an important study on the prediction of disability pension due to depression by effort-reward imbalance at work, mean scores of ERI data were computed for work units. "Thus, all employees in the same work unit were given the same work unit-level score, regardless of

their …own survey responses" (Juvani et al. (2014), 268). The results of this study demonstrated that "high ERI was associated with an increased risk of disability pension due to depression. This association was observed using both work unit- and individual-level measurement of ERI" (Juvani et al. (2014), 270).

Third, as mentioned earlier in this chapter, there is always a need to *examine systematic bias* inherent in subjectively reported data where possible. For instance, several investigations documenting associations of work stress in terms of the ERI model with health included measures of personality traits, e.g. temperament (Tei-Tominaga et al. 2009) and morningness-eveningness (Willis et al. 2008), or of personal response styles, e.g. negative affectivity (Bosma et al. 1998; Ostry et al. 2003; von dem Knesebeck et al. 2009), and adjusted the final estimates for these confounding influences. In future research these arguments concerning the nature of subjective information deserve continuous careful attention.

Innovative developments of the content of the model's indicators and of the format of presenting items or questions are *desirable* as well. For instance, the 'effort' component is narrowly defined and does not adequately capture more recent developments of job demands, e.g. in the human service and IT sectors (Van Vegchel et al. 2001). Including *respondents*' *preferences* concerning inequity or loss aversion and fairness might add a further element to the interpretation of consequences of failed social reciprocity (Fehr and Gintis 2007). With respect to the intrinsic component of the model the focus on over-commitment represents effort- related coping rather than reward-related coping, suggesting that the inclusion of a measure of *reward sensitivity* could improve our understanding of the dynamics of imbalance (Allisey et al. 2012). Of course, there must be a trade-off between the benefits of extending the measurement and the costs of changing established procedures. Methodological innovations are proposed as well. For instance, *vignette questions* representing failed reciprocity in work-related exchange could contribute to a better estimate of the respondents' standards of comparison (Kristensen and Johansson 2008). One could even think of including a short version of an *economic game* into a computer based interview (Dohmen et al. 2011).

In *conclusion*, research on work stress and health based on the ERI questionnaire has produced a substantial amount of new knowledge with relevance to science and policy. At the same time, some methodological problems have been identified, and solutions for further improvement in conjunction with distinct innovations are expected to strengthen this line of inquiry in the near future.

References

Adams, J. S. (1965) Inequity in social exchange. In B. Leonard (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). New York: Academic Press. doi:10.1016/S0065-2601(08)60108-2

Allisey, A., Rodwell, J., & Noblet, A. (2012). Personality and the effort-reward imbalance model of stress: Individual differences in reward sensitivity. *Work and Stress, 26*, 230–251.

Almadi, T., Cathers, I., & Chow, C. M. (2013). An arabic version of the effort-reward imbalance questionnaire: Translation and validation study. *Psychological Reports, 113*, 275–290.

Bagaajav, A., Myagmarjav, S., Nanjid, K., Otgon, S., & Chae, Y. M. (2011). Burnout and job stress among mongolian doctors and nurses. *Industrial Health, 49*, 582–588.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*, 425–440. doi:10.1007/s11336-006-1447-6.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychology Review, 111*, 1061–1071. doi:10.1037/0033-295x.111.4.1061.

Bosma, H., Peter, R., Siegrist, J., & Marmot, M. (1998). Two alternative job stress models and the risk of coronary heart disease. *The American Journal of Public Health, 88*, 68–74.

Buapetch, A., Lagampan, S., Faucett, J., & Kalampakorn, S. (2008). The Thai version of Effort-Reward Imbalance Questionnaire (Thai ERIQ): A study of psychometric properties in garment workers. *Journal of Occupational Health, 50*, 480–491.

Burke, R. (2010). Workplace stress and well-being across cultures: Research and practice. *Cross Cultural Management, 17*, 5–9. doi:10.1108/13527601011016871.

Choi, B., et al. (2009). Cross-language differential item functioning of the job content questionnaire among European countries: The JACE study. *International Journal of Behavioral Medicine, 16*, 136–147. doi:10.1007/s12529-009-9048-2.

de Jonge, J., van der Linden, S., Schaufeli, W., Peter, R., & Siegrist, J. (2008). Factorial invariance and stability of the effort-reward imbalance scales: A longitudinal analysis of two samples with different time lags. *International Journal of Behavioral Medicine, 15*, 62–72.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203–1218.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9*, 522–550.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*, 370–388.

Eum, K.-D., Li, J., Lee, H.-E., Kim, S.-S., Paek, D., Siegrist, J., & Cho, S.-I. (2007). Psychometric properties of the Korean version of the effort-reward imbalance questionnaire: A study in a petrochemical company. *International Archives of Occupational and Environmental Health, 80*, 653–661.

Fayers, P. M., Hand, D. J., Bjordal, K., & Groenvold, M. (1997). Causal indicators in quality of life research. *Quality of Life Research, 6*, 393–406. doi:10.1023/A:1018491512095.

Fehr, E., & Gintis, H. (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology, 33*, 43–64.

Fekete, C., Wahrendorf, M., Reinhardt, J. D., Post, M. W. M., & Siegrist, J. (2014). Work stress and quality of life in persons with disabilities from four European countries: The case of spinal cord injury. *Quality of Life Research, 23*, 1661–1671.

Gamage, A. U., & Seneviratne, R. A. (2015) *Perceived job stress and presence of hypertension among administrative officers in Sri Lanka*. Asia Pac J Public Health [Epub ahead of print]. doi:10.1177/1010539515598834

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161–178.

Griep, R. H., Rotenberg, L., Vasconcellos, A. G. G., Landsbergis, P., Comaru, C. M., & Alves, M. G. M. (2009). The psychometric properties of demand-control and effort-reward imbalance scales among Brazilian nurses. *International Archives of Occupational and Environmental Health, 82*, 1163–1172.

Hadzibajramovic, E., Ahlborg, G., & Grimby-Ekman, A. (2015). *A longitudinal study of the impact of psychoscoial job stressors on symptoms of burnour, synchronous and delayed effects*. Submitted.

Hanson, E. K., Schaufeli, W., Vrijkotte, T., Plomp, N. H., & Godaert, G. L. (2000). The validity and reliability of the Dutch effort-reward imbalance questionnaire. *Journal of Occupational Health Psychology, 5*, 142–155.

Harkness, J., et al. (Eds.). (2010). *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken: Wiley.

Hasselhorn, H., Tackenberg, P., Peter, R., & Group NS. (2004). Effort-reward imbalance among nurses in stable countries and in countries in transition. *International Journal of Occupational and Environmental Health, 10*, 401–408.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology, 12*, 159. doi:10.1186/1471-2288-12-159.

Hernandez Lallement, J., Kuss, K., Trautner, P., Weber, B., Falk, A., & Fliessbach, K. (2013). Effort increases sensitivity to reward and loss magnitude in the human brain. *Social Cognitive and Affective Neuroscience, 9*, 342–349.

Hintsanen, M., Elovainio, M., Puttonen, S., Kivimäki, M., Koskinen, T., Raitakari, O., & Keltikangas-Järvinen, L. (2007). Effort-reward imbalance, heart rate, and heart rate variability: The cardiovascular risk in young finns study. *International Journal of Behavioral Medicine, 14*, 202–212.

Hobfoll, S. E. (1989). Conservation of resources. A new attempt at conceptualizing stress. *The American Psychologist, 44*, 513–524.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. London: Sage.

Holland, P., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale: Lawrence Erlbaum.

House, R., Hanges, P., Javidan, M., Dorfman, P., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage Publication.

Johnston, J. M. (2006). Using computerized ambulatory diaries for the assessement of job characteristics and work-related stress in nurses. *Work and Stress, 20*, 163–172.

Johnston, D. W., Jones, M. C., Charles, K., McCann, S. K., & McKee, L. (2013). Stress in nurses: Stress-related affect and its determinants examined over the nursing day. *Annals of Behavioral Medicine, 45*, 348–356. doi:10.1007/s12160-012-9458-2.

Juarez-Garcia, A., Vera-Calzaretta, A., Blanco-Gomez, G., Gomez-Ortiz, V., Hernandez-Mendoza, E., Jacinto-Ubillus, J., & Choi, B. (2015). Validity of the effort/reward imbalance questionnaire in health professionals from six Latin-American countries. *American Journal of Industrial Medicine, 58*, 636–649. doi:10.1002/ajim.22432.

Juvani, A., Oksanen, T., Salo, P., Virtanen, M., Kivimaki, M., Pentti, J., & Vahtera, J. (2014). Effort-reward imbalance as a risk factor for disability pension: The Finnish Public Sector Study. *Scandinavian Journal of Work, Environment & Health, 40*, 266–277.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 45*, 1623–1630.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*, 1776–1780. doi:10.1126/science.1103572.

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

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *The American Political Science Review, 98*, 191–207.

Kinnunen, A., Feldt, T., & Mäkikangas, A. (2007). Testing the effort-reward imbalance model among Finnish managers: The role of perceived organizational support. *Journal of Occupational Health Psychology, 13*, 114–127.

Kivimäki, M., Vahtera, J., Elovainio, M., Virtanen, M., & Siegrist, J. (2007). Effort-reward imbalance, procedural injustice and relational injustice as psychosocial predictors of health: Complementary or redundant models? *Occupational and Environmental Medicine, 64*, 659–665.

Kompier, M. (2005). Assessing the psychosocial work environment-"subjective" versus "objective" measurement. *Scandinavian Journal of Work, Environment and Health, 31*, 405–408.

Krause, N., Rugulies, R., & Maslach, C. (2010). Effort-reward imbalance at work and self-rated health of Las Vegas hotel room cleaners. *American Journal of Industrial Medicine, 53*, 372–386.

Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics, 15*, 96–117.

Kurioka, S., Inoue, A., & Tsutsumi, A. (2013). Optimum cut-off point of the Japanese short version of the effort-reward imbalance questionnaire. *Journal of Occupational Health, 55*, 340–348.

Lau, B. (2008). Effort-reward imbalance and overcommitment in employees in a Norwegian municipality: A cross sectional study. *Journal of Occupational Medicine and Toxicology, 3*, 9.

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.

Ledoux, J. E. (1989). Cognitive-emotional interactions in the brain. *Cognition and Emotion, 3*, 267–289. doi:10.1080/02699938908412709.

Lehr, D., Koch, S., & Hillert, A. (2010). Where is (im)balance? Necessity and construction of evaluated cut-off pointers for effort-reward imbalance and overcommitment. *Journal of Occupational and Organizational Psychology, 83*, 251–261.

Leineweber, C., Wege, N., Westerlund, H., Theorell, T., Wahrendorf, M., & Siegrist, J. (2010). How valid is a short measure of effort-reward imbalance at work? A replication study from Sweden. *Occupational and Environmental Medicine, 67*, 526–531. doi:10.1136/oem.2009.050930.

Li, J., Yang, W., Cheng, Y., Siegrist, J., & Cho, S.-I. (2005). Effort-reward imbalance at work and job dissatisfaction in Chinese healthcare workers: A validation study. *International Archives of Occupational and Environmental Health, 78*, 198–204.

Li, J., et al. (2012a). Psychometric properties and differential explanation of a short measure of effort-reward imbalance at work: A study of industrial workers in Germany. *American Journal of Industrial Medicine, 55*, 808–815.

Li, J., Loerbroks, A., Shang, L., Wege, N., Wahrendorf, M., & Siegrist, J. (2012b). Validation of a short measure of effort-reward imbalance in the workplace: Evidence from China. *Journal of Occupational Health, 54*, 427–433.

Macias Robles, M. D., Fernandez-Lopez, J. A., Hernandez-Mejia, R., Cueto-Espinar, A., Rancano, I., & Siegrist, J. (2003). Measuring psychosocial stress at work in Spanish hospital's personnel. Psychometric properties of the Spanish version of Effort-Reward Imbalance model. *Medicina Clinica Barcelona, 120*, 652–657.

McGillis Hall, L., & Kiesners, D. (2005). A narrative approach to understanding the nursing work environment in Canada. *Social Science and Medicine, 61*, 2482–2491. doi:10.1016/j.socscimed.2005.05.002.

Msaouel, P., et al. (2012). The effort-reward imbalance questionnaire in Greek: Translation, validation and psychometric properties in health professionals. *Journal of Occupational Health, 54*, 119–130.

Niedhammer, I., Teck, M. L., Starke, D., & Siegrist, J. (2004). Effort-reward imbalance model and self reported health: Cross-sectional and prospective results from the GAZEL Cohort. *Social Science and Medicine, 58*, 1531–1541.

Ostry, A. S., Kelly, S., Demers, P. A., Mustard, C., & Hertzman, C. (2003). A comparison between the effort-reward imbalance and demand control models. *BMC Public Health, 3*, 10.

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.

Pikhart, H., Bobak, M., Siegrist, J., Pajak, A., et al. (2001). Psychosocial work characteristics and self-rated health in four post-communist countries. *Journal of Epidemiology and Community Health, 55*, 624–630.

Pikhart, H., et al. (2004). Psychosocial factors at work and depression in three countries of Central and Eastern Europe. *Social Science and Medicine, 58*, 1475–1482.

Pitts, C. (2014). *Health and wellbeing of older workers: relationships with gender, effort-reward imbalance and flexibility*. Dissertation, Faculty of Health Sciences, University of Sidney.

Rantanen, J., Feldt, T., Hyvönen, K., Kinnunen, U., & Mäkikangas, A. (2012). Factorial validity of the effort-reward imbalance scale: Evidence from multi-sample and three-wave follow-up studies. *International Archives of Occupational and Environmental Health, 86*, 645–656.

Rau, R., Gebele, N., Morling, K., & Rösler, U. (2010). *Untersuchung arbeitsbedingter Ursachen für das Auftreten von depressiven Störungen [An investigation of the occupational causes associated with the incidence of depressive disorders]*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Rödel, A., Siegrist, J., Hessel, A., & Brähler, E. (2004). Fragebogen zur Messung beruflicher Gratifikationskrisen. Psychometrische Testung an einer repräsentativen deutschen Stichprobe. *Zeitschrift für Differentielle und Diagnostische Psychologie, 25*, 227–238.

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine, 25*, 127–141.

Siegrist, J. (1996a). Adverse health effects of high effort – Low reward conditions at work. *Journal of Occupational Health Psychology, 1*, 27–43.

Siegrist, J. (1996b). *Soziale krisen und Gesundheit: Eine Theorie der Gesundheitsförderung am Beispiel von Herz-Kreislauf-Risiken im Erwerbsleben*. Göttingen: Hogrefe.

Siegrist, J. (2002). Effort-reward imbalance at work and health. In P. Perrewé & D. Ganster (Eds.), *Historical and current perspectives on stress and health* (pp. 261–291). Amsterdam: Elsevier.

Siegrist, J., Matschinger, H., Cremer, P., & Seidel, D. (1988). Atherogenic risk in men suffering from occupational stress. *Atherosclerosis, 69*, 211–218.

Siegrist, J., Starke, D., Chandola, T., Godin, I., Marmot, M., Niedhammer, I., & Peter, R. (2004). The measurement of Effort-Reward Imbalance at work: European comparisons. *Social Science and Medicine, 58*, 1483–1499.

Siegrist, J., Wahrendorf, M., von dem Knesebeck, O., Jurges, H., & Borsch-Supan, A. (2007). Quality of work, well-being, and intended early retirement of older employees: Baseline results from the SHARE Study. *European Journal of Public Health, 17*, 62–68. doi:10.1093/eurpub/ckl084.

Siegrist, J., Wege, N., Pühlhofer, F., & Wahrendorf, M. (2009). A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *International Archives of Occupational and Environmental Health, 82*, 1005–1013.

Siegrist, J., Lunau, T., Wahrendorf, M., & Dragano, N. (2012). Depressive symptoms and psychosocial stress at work among older employees in three continents. *Globalization and Health, 8*, 27.

Siegrist, J., Li, J., & Montano, D. (2014). *Psychometric properties of the effort-reward imbalance questionnaire*. www.uniklinik-duesseldorf.de/med-soziologie. Accessed 4 Aug 2015.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers*. San Francisco: Josey Bass.

Tei-Tominaga, M., Akiyama, T., Miyake, Y., & Sakai, Y. (2009). The relationship between temperament, job stress and overcommitment: A cross-sectional study using the TEMPS-A and a scale of ERI. *Industrial Health, 47*, 509–517.

Thomas, W., & Thomas, D. (1928). *The child in America: Behavior problems and programs*. New York: Knopf.

Törnroos, M., et al. (2014). Longitudinal measurement invariance of the effort-reward imbalance scales in the Young Finns study. *Occupational and Environmental Medicine, 71*, 289–294.

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*, 1089–1091. doi:10.1038/nature08785.

Tsutsumi, A., Ishitake, T., Peter, R., Siegrist, J., & Matoba, T. (2001). The Japanese version of the effort-reward-imbalance questionnaire: A study in dental technicians. *Work and Stress, 15*, 86–96.

Tsutsumi, A., Nagami, M., Morimoto, K., & Matoba, T. (2002). Responsiveness of measures in the effort-reward imbalance questionnaire to organizational changes: A validation study. *Journal of Psychosomatic Research, 52*, 249–256.

Tsutsumi, A., Iwata, N., Wakita, T., Kumagai, R., Noguchi, H., & Kawakami, N. (2008). Improving the measurement accuracy of the effort-reward imbalance scales. *International Journal of Behavioral Medicine, 15*, 109–119. doi:10.1080/10705500801929718.

Tsutsumi, A., et al. (2009). Application of item response theory to achieve cross-cultural comparability of occupational stress measurement. *International Journal of Methods Psychiatry, 18*, 58–67.

Van Vegchel, N., de Jonge, J., & Meijer, T. (2001). Different effort constructs and effort-reward imbalance: Effects on employee well-being in ancillary health care workers. *Journal of Advanced Nursing, 34*, 128–136.

van Vegchel, N., de Jonge, J., Bosma, H., & Schaufeli, W. (2005). Reviewing the effort-reward imbalance model: Drawing up the balance of 45 empirical studies. *Social Science and Medicine, 60*, 1117–1131.

von dem Knesebeck, O., Dragano, N., Moebus, S., Jöckel, K.-H., Erbel, R., & Siegrist, J. (2009). Psychosoziale Belastungen in sozialen Beziehungen und gesundheitliche Einschränkungen. *Psychotherapie Psychosomatik Medizinische Psychologie, 59*, 186–193.

Wahrendorf, M., & Siegrist, J. (2014). Proximal and distal determinants of stressful work: Framework and analysis of retrospective European data. *BMC Public Health, 14*, 849.

Wege, N., Dragano, N., Moebus, S., Stang, A., Erbel, R., Jöckel, K. H., & Siegrist, J. (2008). When does work stress hurt? Testing the interaction with socioeconomic status in the Heinz Nixdorf Recall Study. *The Journal of Epidemiology and Community Health, 62*, 338–341.

Weyers, S., Peter, R., Boggild, H., Jeppesen, H., & Siegrist, J. (2006). Psychosocial work stress is associated with poor self-rated health in Danish nurses: A test of the effort-reward imbalance model. *Scandinavian Journal of Caring Sciences, 20*, 26–34.

Willis, T. A., O'Connor, D., & Smith, L. (2008). Investigating effort-reward imbalance and work-family conflict in relation to morningness-eveningess and shift work. *Work and Stress, 22*, 125–137. doi:10.1080/02678370802180558.

Yadegarfar, G., et al. (2013). Psychometric properties of the Farsi version of effort-reward imbalance questionnaire: A longitudinal study in employees of a synthetic fibre factory in Iran. *International Archives of Occupational and Environmental Health, 86*, 147–155. doi:10.1007/s00420-012-0750-z.

Yang, Z., Yang, B., & Li, J. (2015). Perspectives on compensation and legislation of death due to work overload-karoshi. *QJM, 108*, 349–350. doi:10.1093/qjmed/hcu207.

Yokoyama, K., Hirao, T., Yoda, T., Yoshioka, A., & Shirakami, G. (2014). Effort-reward imbalance and low back pain among eldercare workers in nursing homes: A cross-sectional study in Kagawa Prefecture, Japan. *Journal of Occupational Health, 56*, 197–204.

Zurlo, M., Pes, D., & Siegrist, J. (2010). Validity and reliability of the effort-reward imbalance questionnaire in a sample of 673 Italian teachers. *International Archives of Occupational and Environmental Health, 83*, 665–674.