cannot always be comprehensive, especially in repeated follow-up studies. Valid single-item methods would offer a means with which to report the results of interesting intervention trials dealing with work and worker well-being when comprehensive scales are not accepted by the target organization. The Occupational Stress Questionnaire (17, 18) was developed for this purpose according to the single-item principle for use by occupational health personnel in monitoring perceived well-being and the psychosocial factors related to it. The questionnaire is based on the psychological theory of work stress (6, 19), and it also includes the dimensions of job demands, control, and social support (1).

The single-item measure of stress symptoms included in the Occupational Stress Questionnaire was developed in the beginning of the 1970s on the basis of both symptom checklists used in mental health screening and clinical experience with normal patients in occupational health settings. The question refers to the general experience of stress, not to work-related stress, as follows: "Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because his/her mind is troubled all the time. Do you feel this kind of stress these days?" The response is recorded on a 5-point Likert scale varying from "not at all" to "very much". The question is used in individual and group screening in occupational health services, in organizational assessment (18), and in population studies (20). It was used as a national indicator of psychosocial harm in describing national profiles as a response to the European Office of the World Health Organization in developing criteria for auditing workplace health systems (21). Such long-term use of the question in different contexts shows a priori or face validity that can be interpreted as an intuitive estimate of content validity (22).

A measure is valid when it measures what it is purported to measure. According to the principles of psychological test validation, "validity" refers to the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores, and test validation is the process of accumulating evidence to support such inferences (23). The conceptualization of validity varies somewhat in the fields of psychology, epidemiology, and sociology, but common to these traditions is the distinction between construct validity and criterion validity. Construct validity refers to the degree to which the measure captures the hypothetical quality or trait (ie, the construct). "The estimate of construct validity is always changing with the accumulation of further evidence about the traits and qualities that underlie the construct [p 781]" (22). Criterion validity can be established in relation to an independent validated criterion method that is concurrently available with the investigated method or in relation to a future outcome (22, 24, 25). Measurement reliability is a prerequisite for the empirical testing of validity.

In psychology, content validity is also emphasized, as it is easier to study it empirically than it is to study construct validity, through convergence and divergence with other measures. On the whole, investigating validity does not deviate from the general scientific procedures used to confirm theories (26, 27). Empirical testing of validity is not always possible, but almost any information gathered in the process of developing or using a test (or method) is relevant to its validity (28).

The concurrent validity of a method can be investigated by comparing its results with those of a method with well-characterized properties. Generally, factor analyses of construct and content validity, especially structural equation modeling, are used for this purpose. For single items, traditional methods are available for investigating validity. The multitrait-multimethod matrix, correlations with variables assumed to measure the same concept, experimental designs, and investigations of response processes are some examples (29). Longitudinal data would offer the best means of estimating construct validity and predictive criterion validity, but the numerous problems in implementing longitudinal study designs in worklife limit the applicability of this approach (30, 31). Especially in the context of method development, organizations and employees are reluctant to spend their time responding to extensive test batteries and undergoing repeated measurements. Frequent changes in modern organizations also limit the possibilities to carry out randomized reference studies and interpret the changes observed.

Borg et al (32) reported positive experience with the predictive validity of a single-item measure of self-rated health. Their results showed that, in a 5-year follow-up of a working population, repetitive work, psychological demands, low social support, job insecurity, and ergonomic exposures were significant predictors of the worsening of self-rated health. Wanous et al (33) carried out a meta-analysis on the validity of single-item measures of job satisfaction. According to their results, it is acceptably reliable and valid to use a single-item measure for a concept such as job satisfaction, which is located between factual questions and more abstract or vague psychological concepts. According to their meta-analysis, the estimated lower limit of the reliability of single-item measures of job satisfaction is 0.67. Although Wanous et al (33) recommended the use of sum scales whenever possible, they listed certain, often practical reasons for using single-item measures. Reduced costs, increased face validity for the respondent, and problems related to the construction of sum scales support the use of single items instead of sum scales. Item bias, for example, the blurring or reversal of information, has been shown to be common with sum scales measuring work characteristics (34, 35).