

## Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements

**Dan Hasson, RN, PhD; Bengt B. Arnetz, MD, PhD**

Authors are affiliated with Uppsala University/CEOS, Department of health and Caring Sciences, Section for Social Medicine. **Contact Author:** Dan Hasson, Department of Public Health and Caring Science, Section for Social Medicine, Uppsala University/CEOS, Uppsala Science Park, SE-751 85 Uppsala, Sweden; phone: +46 70 728 42 98; fax: +46-18-506404; email: [dan.hasson@pubcare.uu.se](mailto:dan.hasson@pubcare.uu.se)

Submitted July 19, 2005; Revised and Accepted October 1, 2005

---

### **Abstract**

*Context: Psychosocial exposures commonly show large variation over time and are usually assessed using multi-item Likert indices. A construct requiring a five-item Likert index could possibly be replaced by a single visual analogue scale (VAS). Objective: To: a) evaluate validity and relative reliability of a single VAS compared to previously validated Likert based items and indices measuring the same construct, b) detect possible statistically significant differences in absolute levels between the single VAS and Likert items and indices respectively. Design: Cross-sectional study conducted in May 2004. Methods: 805 participants responded to a web-based questionnaire including both VAS and Likert based items. Intraclass correlations were utilized to assess agreement between VAS and Likert scales/indices and Marginal homogeneity tests were utilized to detect possible differences in absolute levels. Results: Moderate to strong correlations were found in responses between VAS and Likert based items and indices, and significant differences in absolute levels in seven out of eleven scales. Conclusion: Single VAS questions can, in some cases of uniform construct, replace a single Likert item and also be comparable, but not interchangeable, with multi-item Likert indices.*

**Key Words:** Comparison, health, Likert, measurement, psychosocial, stress, validation, VAS

---

## Introduction

There is now extensive evidence that psychosocial factors may contribute to the development of ill health<sup>1-9</sup>. Psychosocial exposures, such as, work-strain, family conflicts and even socioeconomic stressors commonly show large individual variations over time<sup>10-12</sup>. To further explain how stress is associated with health, it is proposed that we need to develop measurement methods that can be repeated regularly and that are sensitive to variations in exposures and outcomes over time. Moreover, for a measurement method to be meaningful and useful it must be shown that it is reliable (i.e. accurate and consistent, e.g. measure similar levels in stable subjects), and valid (i.e. if it really measures what it intends to, e.g. a health-related questionnaires' ability to actually assess health)<sup>13,14</sup>. If change is to be measured over time another property, responsiveness, is of importance. The responsiveness refers to the ability of an instrument to detect clinically significant (as distinct from statistically significant) changes over time, even if those changes are small<sup>15,16</sup>.

Two commonly utilized methods of psychosocial measurement are the Likert scale and the Visual Analogue Scale (VAS). The Likert scale<sup>17</sup> is the most widely used scaling technique<sup>13</sup> and commonly used in various stress and health research studies<sup>18</sup>. These scales typically consist of items that for example require respondents to rate their degrees of agreeing or disagreeing with various declarative statements. Usually three to seven response alternatives are used, but there are different opinions about the optimal number of response alternatives<sup>13,18-20</sup>. In any case, wording of the response alternatives most probably affect the responses<sup>21</sup>. VAS is a simple method for measuring subjective experience. Typically, a VAS consists of a 10 centimeter line anchored at each end by words descriptive of opposing statements or the minimal and maximal extremes of the dimension being measured, e.g. "No pain" and at the other end "Worst possible pain". Respondents are required to place a mark on that line<sup>13,22,23</sup>. VAS has mostly been used in prior studies measuring pain<sup>24</sup>, mood<sup>25,26</sup>, fatigue<sup>27,28</sup>, respiration<sup>28</sup>, functional capacity<sup>29</sup>, tension<sup>2</sup> and in the classification of psychiatric patients<sup>30</sup>.

### *Validity and reliability*

Both Likert and VAS scales have been evaluated in terms of reliability, validity and responsiveness. In general, both scaling methods seem to be both reliable, valid and responsive<sup>15,21-23,31-33</sup>. Even though the use of VAS is often

recommended in applied research, there has been criticism concerning the reliability, validity and interpretation of the study results<sup>34</sup>. Many studies have also compared VAS and Likert scales with regard to reliability and/or validity and/or user-friendliness and/or responsiveness in different settings<sup>15,16,21,22,27,30,32,33,35-49</sup>. The results are contradictory, and whereas most of these studies find significant correlations or no differences between ratings, some find significant differences in ratings between the two types of scales. Du Toit et al.<sup>16</sup> suggest that this difference might occur because the VAS is more sensitive to detect small differences than the Likert scale. However, Smith et al.<sup>37</sup> argue that although changes in their study were statistically significant, they may not represent clinically significant changes.

### *Advantages and disadvantages with VAS and Likert scales*

Both VAS and Likert scales have specific advantages and disadvantages. Supporting statements of the Likert scale are that it is easier to use and understand both for the researcher and the respondent and that coding as well as interpretation is easier compared to VAS. It also takes less time to explain to patients<sup>21,33,39,46</sup>. The use of Likert scales have also been found to be easier for young children to understand and answer correctly compared to VAS<sup>45,47</sup>. In addition, Vickers<sup>21</sup> argues that Likert scales are more responsive than VAS. Limitations with the Likert scale is that wording of the descriptive categories most probably affect the responses and artificial categories might not be sufficient to describe a complex continuous, subjective phenomenon<sup>21,22</sup>. Furthermore, too many response categories may lead to difficulties in choosing and too few may not provide enough choice or sensitivity, forcing the respondent to choose an answer that does not represent the person's true intent. Finally, a total score from a multi-item Likert index may be the result of many different combinations of ratings, which leads to a loss of information about the scale items<sup>50</sup>. Moreover, it has been indicated that the use of sum scores may lead to incorrect conclusions<sup>18</sup>.

With regard to VAS, it has been argued that major advantages are that it is relatively easy to use, and to understand, particularly by less educated raters and immigrants<sup>23,47,51,52</sup>. Some researchers claim that it is easier to use than the Likert scale and that VAS is preferred by the raters<sup>16,22,35,40,41</sup>. Furthermore, in contrast to Vickers<sup>21</sup>, it has been suggested that VAS has a better responsiveness (i.e. ability to detect clinically significant change) than the Likert scale and might also be more reliable and valid<sup>15,16,22,23,36,38,40,41,44</sup>. Others indicate that VAS and

Likert scales are comparable with regard to reliability and validity and yield similar results<sup>15,33,47</sup>. A widely cited article by Onhaus and Adler<sup>43</sup> states that VAS seems to assess more closely what patients actually experience. Disadvantages with the VAS are that it is difficult to understand for some users, require a significant time and commitment for instruction and administration, and involves more work than a Likert scale<sup>23,41,53</sup>. It might also not be a valid measure for young children since they do not understand and answer the VAS correctly<sup>45</sup>. Furthermore, it has been suggested that a mark on the VAS have no interpretable meaning<sup>32</sup> and VAS might be less specific and have worse precision than the Likert scale<sup>21</sup>.

It is clear that the studies reporting on the advantages and disadvantages using VAS vs. Likert scales often report contradictory findings. Overall, the results from the above mentioned studies indicate that the context and setting in which the scales are being used seems to be of importance for the reliability, validity and usefulness of a rating scale. Therefore, there is not enough evidence to conclude that one scale type generally is "better" than the other. Rather one scale type might be more suitable for a specific context or setting. Furthermore, weaknesses of previous studies make it difficult to compare findings and draw uniform, reliable and generalizable conclusions. There is, for example, no uniform agreement as to how comparisons between VAS and Likert scales should be conducted. A wide range of statistical techniques and calculations have been used in order to reach conclusions. Some treat the VAS as interval or even ratio data<sup>18,22,52</sup> and consequently use parametric methods, whereas most others regard VAS as ordinal data<sup>18,32,34,54</sup> and therefore use non-parametric methods. Price and associates have in a number of studies<sup>52,55</sup> attempted to provide evidence that VAS is valid as a ratio scale. They demonstrated that VAS could predict experimental pain intensity and stimuli to be twice as intense as a standard stimulus. However, the data neither provided evidence that such relationships existed along the entire length of the VAS, nor that a VAS score of 80 represents twice as much pain as a score of 40. On the other hand, Maxwell<sup>56</sup> has concluded that it generally makes little difference whether parametric or non-parametric test are used to analyze VAS data. Finally, most studies differ in design, include different kinds of populations, often with few participants, and are conducted in a wide range of settings with scales for different purposes of use.

Since several previous studies have found the VAS to be responsive in different settings, it might

be suitable to measure psychosocial exposures on a regular basis, as they may show large variations over time. Long-term assessment of psychosocial exposures could render new insights about the nature of these exposures in general, but also generate knowledge about trends and scorings that might be beneficial or harmful to long-term health. The present study is a pilot-study that aims at validating a number of VAS that will be used in a more extensive trial. Because the validity of a scale is essential, the aims of the present study were to:

- a) Evaluate validity and relative reliability (see method section for definition) of self-ratings on recently constructed single VAS compared to previously thoroughly validated (and a few non-validated) Likert based single items and multiple-item indices measuring the same construct.
- b) Detect possible statistically significant differences in absolute levels between the single VAS and Likert items and indices respectively.

Based on findings in most of the previously reviewed literature in the present study and theory available, it was hypothesized that there would be significant and medium to strong correlations between VAS and Likert items and indices and no statistically significant differences in absolute levels.

## Methods

### *Participants*

Participants were recruited via two web sites; one web site of the centre of environmental ill-health and stress ([www.ceos.nu](http://www.ceos.nu)) at the Academic Hospital in Uppsala and one site offering a stress management and health promotion tool ([www.pql.se](http://www.pql.se)) previously developed in a research project at Uppsala University. The invitation to participate in the study was posted in the news section on the main page of the web sites and interested visitors could click on a link to receive some more information. Furthermore, 3016 randomly selected registered users of the website [www.ceos.nu](http://www.ceos.nu) received an e-mail asking them as to their interest to participate in the present study. The participants were informed that a questionnaire was to be validated and that their participation was voluntary and anonymous. There were approximately 13,400 visits from about 2,700 unique individuals (some individuals visited the sites on a regular basis) at the websites during the two-week duration of the study. 805 individuals chose to participate in the study, out of which

students, unemployed, individuals on sick-leave and pensioners were excluded in order to make the population more homogenous with regard to socioeconomic factors. Thus, the final number of participants was 633. Background characteristics of the participants are presented in Table 1.

### *Questionnaire*

The questionnaire with Likert and VAS based response alternatives consisted of socioeconomic background as well as stress- and health-related questions. The background questions included age (<20, 20-30, 31-45, 46-60, >60 yrs), gender (male vs. female), marital status (married/co-inhabiting/live-apart vs. single), educational level (primary school, high school, academic degree (BS, BA), and higher academic degree), income (<\$12 000, \$12-30 000, \$30-45 000, >\$45 000 pr annum), occupation (working, pensioner, sick-leave, student, unemployed) and financial situation (very poor – very well). The stress- and health-related questions included the single item on self-rated health (SRH)<sup>57-59</sup>, the Karolinska Sleep Questionnaire (KSQ)<sup>60</sup>, the indices mental energy and work-related exhaustion from the Quality Work Competence (QWC) questionnaire<sup>10,61</sup> as well as newly constructed single VAS and Likert based items to be compared. Items, topics and indices covered by the questionnaire are presented in Table 2.

### *Data analysis, reliability and validity*

The statistical program SPSS 13.0 program for PC was used for data analysis. Since VAS and Likert scales are commonly considered to be ordinal in nature, non-parametric tests were utilized. However, as previous studies have indicated that VAS may also be considered as interval or even quote scales, corresponding parametric tests were also conducted<sup>18,22,52,54</sup>. In accordance with Maxwell's<sup>56</sup> findings, there were no differences in the results of the parametric and non-parametric tests in the present study. Hence, the results of the parametric tests are not presented.

To assess the consistency of the Likert based indices, confirmatory factor analysis was conducted. Since covariance between items in the index would probably exist, Principal component analysis with Equamax rotation was used for factor extraction. Internal reliability of the Likert indices was measured using Cronbach's  $\alpha$ .

Since the present study was cross-sectional, the reliability of each response option could not be established. Instead, relative reliability<sup>47</sup>, i.e. the accordance between VAS and Likert scorings, was estimated using Intraclass correlations. Based on results from prior usage of the validated items and

scales, it was assumed that at least one of the scales would represent the "true" answer. Consequently, a high correlation coefficient would indicate a high reliability of both scale types and vice versa.

Criterion-related validity, or more specifically, concurrent validity refers to the degree to which the VAS scores correlated with an external criterion, in this case the previously validated Likert items and indices<sup>13,14</sup>. Most of the studies on the validity of VAS have utilized a criterion-related approach. An established instrument was used as the criterion and correlations have ranged from .42 to .91<sup>23</sup>. The construct validity refers to the degree to which an instrument measures the construct under investigation and ideally requires a pattern of consistent findings from various studies<sup>13,14,62</sup>. The Likert based items and indices used in the present study have shown such qualities. A common approach is to compare different types of scales that aim at measuring the same variable; more specifically, one scale is compared with an accepted standard or "true" state. The level of agreement between the scales reflects the extent to which the scales are interchangeable<sup>34</sup>. The construct validity is then determined by how closely the scales correlate. It was assumed that construct validity would exist in cases where a single VAS was compared with a single Likert item with the same wording of the question and extreme response categories. In cases where a single VAS was compared with a multi-item Likert index, it was assumed that construct validity would exist where the correlations were high. Similarly convergent validity, a form of construct validity, is derived from the correlations between two different methods measuring the same construct (and thus are supposed to converge).

Marginal homogeneity tests were used to test possible statistical differences in absolute levels between VAS and Likert scorings. Before differences in absolute levels were investigated, all scorings from Likert based items and indices were converted into percentages. The formula for this conversion was (Likert scale score or index sum score - min) / (max - min) x 100, where "min" is the lowest score/sum score that a scale/index can assume and "max" is the highest. Like with the VAS, a higher percentage is more beneficial. The VAS scorings already ranged from 0-100. Additionally, to discover potential systematic bias in the location of scorings, i.e. end-aversion bias, percentages of scorings in each category of the Likert scale were compared with the percentage of the VAS. In case of absence of systematic bias, these percentages would have to be comparable. Therefore the VAS were recoded into five even categories (scores 1-20 into category 1, 21-40 into category 2, etc.). Marginal homogeneity tests

were used to test whether the number of respondents scoring in each category was consistent across response options.

The statistical significance level was set at  $p=0.05$  for all analyses. The study was part of a larger trial of a web-based health promotion and stress management program that had been approved by the ethics committee of Uppsala University (Dnr 01-188) and Karolinska Institute (Dnr 01-355).

## Results

In concordance with previous publications on the KSQ, factor analysis resulted in three indices that were derived from the questionnaire, namely: disturbed awakening, disturbed sleep and sleepiness. The index mental energy (a person's mental and cognitive well-being) and the index work-related exhaustion remained unchanged. Factor loadings and Cronbach's  $\alpha$  for the Likert based indices are presented in Table 3. A more extensive table with detailed factor loadings for each index can be obtained from the first author.

The single VAS and single Likert items measuring the same construct were highly correlated, whereas there were moderate to strong correlations between the single VAS and the Likert indices (Table 4). Thus, respondents who scored high on the VAS for the variable "self-rated health" also tended to score high on the corresponding Likert based item. However, in spite of a moderate to strong correlation, there were statistically significant differences in absolute levels in seven out of eleven assessed variables. A lower shared variance ( $r^2$ ) between the two compared scales might explain some of these differences in absolute levels. The Marginal homogeneity tests revealed that the absolute levels differed significantly in 7 out of 11 scales. Correlations and differences in absolute levels are presented in Table 4.

There appeared to be systematic end-aversion bias in scorings on the Likert scales. A larger percentage of the respondents tended not to mark the two extreme ends of the Likert scales as compared with scorings on the corresponding VAS. Thus, on the Likert scales but not the VAS, respondents seemed to mark the three middle response categories more often, suggesting end-aversion bias.

## Discussion

In the present study we evaluated validity and relative reliability of self-ratings on non-validated single VAS-items compared to previously thoroughly validated (and a few non-validated) Likert based single items and indices measuring the same construct. To assess the magnitude of the correlations between the single VAS and corresponding Likert scales/indices we adopted the Cohen criteria, previously used by Van Dijk et al.<sup>31</sup>. Thus, a correlation coefficient of 0.10-0.29 is considered weak, 0.30-0.49 moderate and  $r = 0.50$  strong. Consequently, we found significant, one moderate, and otherwise strong correlations between VAS and Likert items and indices respectively, ranging from 0.44-0.94 ( $p<.001$ ). For example, high self-ratings of health on a VAS corresponded and correlated significantly and strongly with high self-ratings on the matching Likert based item or index. This finding is in line with several previous studies that have found moderate to strong correlations between VAS and Likert scales<sup>21, 27, 30, 33, 36, 42, 46-48</sup>. This suggests that both scale types are comparable with regard to (relative) reliability. It also indicates that VAS may be valid measures, since a strong correlation between related constructs can be assumed to be a sign of criterion-related (concurrent) and construct-related (convergent) validity. This assumption is especially true for the single VAS, which showed strong correlations ( $r = .90-.94$ ,  $p<.001$ ) with the Likert counterpart.

Correlations were less strong where scorings on a single VAS were compared with scorings on a multi-item Likert based index ( $r = .44-.83$ ,  $p<.001$ ). There could be several reasons for this, and one explanation could be that the scales might be interpreted in different ways. For example, VAS can be described as a uni-dimensional model that measures a construct, e.g. mental energy, with one item. Multi-item Likert indices, however, measure the same construct using multiple items. If these items cover different aspects of the same variable the index is considered to be uni-dimensional, and otherwise multidimensional<sup>18</sup>. Thus, if the construct is multidimensional with contradicting aspects, one VAS might not have the ability to capture it. A sum score from a Likert index however, might not be informative, and it has been reported that a low item score overweighs high score on other items, and so the use of sum scores may lead to incorrect conclusions<sup>18</sup>. Furthermore, the opposing statements of VAS represent the two extremes of a variable, which might capture aspects that are beyond the reach of the Likert index. In such case

VAS would render more variation in responses and consequently be a better and more accurate measure of the variable. It has been proposed that a single-item VAS is more appropriate in measuring both uni-dimensional and multi-dimensional constructs<sup>49</sup>. Svensson<sup>18</sup> poses a general statement that: "...One way to avoid aggregation of multi-item scales is to construct a global single item measure. Such an approach might be more valid than multi-dimensional multi-item instruments..." Others suggest that single-item VAS should only be used for uni-dimensional constructs and not multi-dimensional<sup>23</sup>. Considering these views, it is difficult to conclude that Likert indices would be more relevant for multidimensional concepts as compared to a single VAS. Furthermore, such a complex construct as health has been shown to be reliably assessed by the discrete single item self-rated health<sup>57-59</sup>. Additionally, if a concept is multidimensional, confirmatory factor analysis would identify not one but multiple components/factors. This would still result in more than one index and thus offer no advantage over VAS.

For exposures that typically exhibit large variation over time and therefore could benefit from repeated measures, e.g. daily stressors, a VAS appears to be more appropriate than a Likert index when regarding results from previous research. A single-item VAS is less time consuming compared to a multi-item index and VAS most likely also more responsive<sup>15, 16, 22, 23, 36, 41, 44</sup>. Taken together, the findings of the present and previous studies suggest that a single-item VAS might be more or less comparable and interchangeable with a single-item Likert question. However, a single VAS might be associated and perhaps comparable with a multi-item Likert index, but not interchangeable.

To our knowledge, studies to date have only occasionally investigated both correlations and absolute levels when comparing VAS with Likert based scales. In order to compare results within and between studies it is not sufficient to show correlations between the two types of scales. It is also important to show that the two scales exhibit similar calibrations, i.e., show similar absolute scores. Even though there were moderate to strong correlations between VAS and Likert scales in the present study, the absolute levels differed significantly in 7 out of 11 scales with the Marginal homogeneity tests. This result was similar for all question areas except for the single VAS comparing mental energy last month with the corresponding Likert based index. These findings indicate that there are significant similarities in

answering patterns on different scale types, but most often the absolute levels differ. This might be an important issue to consider if absolute levels from previous measurement methods are set as delimitations for, as an example, unhealthy levels of some variable.

A possible explanation for the differences in absolute scores may be the end-aversion bias found in the Likert scales compared to the VAS, rendering a difference in scores. On the Likert scales, a higher percentage of the respondents tended to mark the middle options and avoided the extreme ends of the scales. Additionally, a lower shared variance ( $r^2$ ) between the two compared scales might also explain some of the differences in absolute levels.

In order to improve knowledge about the role of psychosocial exposures that commonly fluctuate over time, in the development and courses of disease it is important to develop sensitive and dynamic exposure measurement methods. Previous studies report that VAS are easily administered and suitable for frequent assessments<sup>15, 16, 22, 23, 36, 41, 44</sup>. However, not many studies have compared the validity and reliability of VAS in comparison to Likert items and indices in the arena of psychosocial health research.

#### *Considerations with online measurements*

The present study was conducted via an online questionnaire. This fact poses a question about the generalizability of the present results to paper-based questionnaires and other online surveys. In a study by Riva et al.<sup>63</sup> an online and offline version of a Likert based questionnaire on Internet attitudes and computer use was compared. The online questionnaire was found to have good test-retest reliability. Furthermore, it was reported that online and offline scorings on questionnaires were similar, but not identical. For example, some online subscales loaded on other items than offline ones. Therefore reassessment of validity was recommended for questionnaires to be converted for online use. However, online data collection neither statistically enhanced nor diminished the consistency of responses, nor compromised the integrity of the test. Andersson et al.<sup>64</sup> also found that online and offline questionnaires yielded comparable results in terms of psychometric properties, but indicated a slight difference in scorings. The prevalence of depression and anxiety was marginally higher in online questionnaires compared with offline. They concluded that the online questionnaire resulted in valid data consistent with previous research. The present study also yielded similar or higher individual

factor loadings and Cronbach's  $\alpha$  compared to previous offline assessments of the QWC-questionnaire and the KSQ. For instance, the offline QWC-questionnaire indices have exhibited individual factor loadings of 0.5 or higher and Cronbach's  $\alpha > 0.7$ <sup>61</sup> and KSQ indices have exhibited similar factor loadings, e.g. "disturbed sleep" 0.72 with Cronbach's  $\alpha > 0.79$ <sup>60</sup>.

### Weaknesses

There were some weaknesses in this study. Firstly, when using an online questionnaire there are numerous possible temporary factors, including technical issues, which might occur and affect the test results. However, since this technology has been used and refined on thousands of individuals during the past three years, all of the major issues have been solved. Secondly, since the present study was cross-sectional only relative reliability could be computed, as test-retest procedures could not be conducted. Thirdly, it can be discussed whether or not correlations between scales were influenced by each other, as they were both scored in the same questionnaire during the same time period. However, it has previously been shown that correlation is maintained irrespective of if the scoring of the scales is separated in time or not<sup>65</sup>. Fourthly, self-selection bias needs to be considered since the individuals that volunteered to participate in this kind of study and via the Internet might not be representative of the general population. Furthermore, skewness in demographics, e.g. 72% of the respondents were female and 8% were under 30 years old, etc., may also decrease the generalizability of the results to the general population. Finally, we have no other information on the respondents apart from the areas covered by the background questions.

### Conclusions

Overall, there is no conclusive evidence that neither VAS nor Likert based scales are superior to one another from a statistical point of view<sup>40</sup>. Rather, the context of application and circumstances of use seems to be of greater importance<sup>34</sup>. The results of the present study imply a similarity in response behavior between VAS and Likert scales. The advocates for VAS claim that it is more responsive, i.e. exact in detecting small clinically significant changes, and hence also more reliable and valid. However, there is no uniform agreement whether that is the case or not. Those that argue for the use of Likert scales claim that it is easier to administer for the researcher as well as the respondent. Also in this

case there is no uniform agreement and with the advance of information technology this is no longer the point. An interesting aspect that supports good overall validity for both scales, is that both VAS and Likert scales have shown comparably strong correlations with physiological markers<sup>10,38</sup>.

Altogether, the results of the present and previous studies comparing VAS and Likert scales imply that both types of scales are relevant for the measurement of fluctuating variables, such as psychosocial exposures. In the present study, correlations between VAS and Likert based items and indices were moderate to strong, but the absolute levels did differ in over half of the assessed scales. More research is needed to establish weaknesses and strengths of each scale type in different contexts and in relation to different exposures of interest. Single VAS questions can, in some cases of uniform construct, replace a single Likert item and also be comparable, but not interchangeable, with multi-item Likert indices. This is especially relevant when assessing fluctuating variables that preferably should be measured repeatedly over time since retrospective self-ratings might not be representative assessments of a longer time period<sup>11,66</sup>. Moreover, it would be facilitating and time-saving for the researcher as well as the respondents and most likely enhance our ability to better understand the relationships between psychosocial exposures of every day life, biological mechanisms and health.

## REFERENCES

1. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol.* 1979;109(2):186-204.
2. Holte KA, Vasseljen O, Westgaard RH. Exploring perceived tension as a response to psychosocial work stress. *Scand J Work Environ Health.* Apr 2003;29(2):124-133.
3. Fink G. *Encyclopedia of stress.* London: Academic Press; 2000.
4. Jenkins CD. Recent evidence supporting psychologic and social risk factors for coronary disease. *N Engl J Med.* 1976;294(19):1033-1038.
5. Karasek R, Baker D, Marxer F, Ahlbom A, Theorell T. Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. *Am J Public Health.* 1981;71(7):694-705.

6. Knox SS. Psychosocial Stress and the Physiology of Atherosclerosis. In: Theorell T, ed. *Everyday biological stress mechanisms*. Basel: Karger; 2001:139-151.
7. Marmot MG, Syme SL. Acculturation and coronary heart disease in Japanese-Americans. *Am J Epidemiol*. 1976;104(3):225-247.
8. Steptoe A, Marmot M. Burden of psychosocial adversity and vulnerability in middle age: associations with biobehavioral risk factors and quality of life. *Psychosom Med*. Nov-Dec 2003;65(6):1029-1037.
9. Westerlund H, Ferrie J, Hagberg J, Jeding K, Oxenstierna G, Theorell T. Workplace expansion, long-term sickness absence, and hospital admission. *Lancet*. Apr 10 2004;363(9416):1193-1197.
10. Arnetz BB. Techno-stress: a prospective psychophysiological study of the impact of a controlled stress-reduction program in advanced telecommunication systems design work. *J Occup Environ Med*. 1996;38(1):53-65.
11. Stone AA, Shiffman S. Reflections on the intensive measurement of stress, coping, and mood, with an emphasis on daily measures. *Psychology and Health*. 1992;7(2):115-129.
12. Snell Dohrenwend B, Dohrenwend BP, eds. *Stressful life events: their nature and effects*. New York: Wiley & Sons Inc.; 1974.
13. Polit DF, Beck CT. *Nursing research: principles and methods*. 7. ed. Philadelphia: Lippincott Williams & Wilkins; 2004.
14. Carmines EG, Zeller RA. *Reliability and validity assessment*. 2. pr. ed. Beverly Hills: Sage; 1979.
15. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. *J Manipulative Physiol Ther*. Jan 1998;21(1):1-7.
16. du Toit R, Pritchard N, Heffernan S, Simpson T, Fonn D. A comparison of three different scales for rating contact lens handling. *Optom Vis Sci*. May 2002;79(5):313-320.
17. Likert R. A technique for the development of attitude scales. *Educational and Psychological Measurement*. 1952;12:313-315.
18. Svensson E. Construction of a single global scale for multi-item assessments of the same variable. *Stat Med*. Dec 30 2001;20(24):3831-3846.
19. Bowling A. *Measuring disease: a review of disease specific quality of life measurement scales*. Buckingham; Philadelphia: Open University Press; 1995.
20. Fayers PM, Jones DR. Measuring and analysing quality of life in cancer clinical trials: a review. *Stat Med*. Oct-Dec 1983;2(4):429-446.
21. Vickers AJ. Comparison of an ordinal and a continuous outcome measure of muscle soreness. *Int J Technol Assess Health Care*. Fall 1999;15(4):709-716.
22. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med*. Nov 1988;18(4):1007-1019.
23. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health*. 1990;13(4):227-236.
24. Feuerstein M, Carter RL, Papciak AS. A prospective analysis of stress and fatigue in recurrent low back pain. *Pain*. 1987;31(3):333-344.
25. Aitken RC. Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine*. 1969;62(10):989-993.
26. Bond A, Lader M. The use of analogue scales in rating subjective feelings. *British Journal of Medical Psychology*. 1974;47:211-218.
27. Brunier G, Graydon J. A comparison of two methods of measuring fatigue in patients on chronic haemodialysis: visual analogue vs Likert scale. *Int J Nurs Stud*. 1996;33(3):338-348.
28. Stark RD, Gambles SA, Lewis JA. Methods to assess breathlessness in healthy subjects: a critical evaluation and application to analyse the acute effects of diazepam and promethazine on breathlessness induced by exercise or by exposure to raised levels of carbon dioxide. *Clin Sci (Lond)*. 1981;61(4):429-439.
29. Scott PJ, Huskisson EC. Measurement of functional capacity with visual analogue scales. *Rheumatol Rehabil*. 1977;16(4):257-259.
30. Remington M, Tyrer PJ, Newson-Smith J, Cicchetti DV. Comparative reliability of categorical and analogue rating scales in the assessment of psychiatric symptomatology. *Psychol Med*. Nov 1979;9(4):765-770.
31. van Dijk M, Koot HM, Saad HH, Tibboel D, Passchier J. Observational visual analog



- scale in pediatric pain assessment: useful tool or good riddance? *Clin J Pain*. Sep-Oct 2002;18(5):310-316.
32. Svensson E. Comparison of the Quality of Assessments Using Continuous and Discrete Ordinal Rating Scales. *Biometrical Journal*. Jul 28 2000;42(4):417-434.
33. Jaeschke R, Singer J, Guyatt GH. A comparison of seven-point and visual analogue scales. Data from a randomized trial. *Control Clin Trials*. Feb 1990;11(1):43-51.
34. Svensson E. Concordance between ratings using different scales for the same variable. *Stat Med*. Dec 30 2000;19(24):3483-3496.
35. Castle NG, Engberg J. Response formats and satisfaction surveys for elders. *Gerontologist*. Jun 2004;44(3):358-367.
36. Seymour RA. The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. *Eur J Clin Pharmacol*. 1982;23(5):441-444.
37. Smith GA, Strausbaugh SD, Harbeck-Weber C, Cohen DM, Shields BJ, Powers JD. New non-cocaine-containing topical anesthetics compared with tetracaine- adrenaline-cocaine during repair of lacerations. *Pediatrics*. 1997;100(5):825-830.
38. Grant S, Aitchison T, Henderson E, et al. A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest*. 1999;116(5):1208-1217.
39. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis*. 1987;40(12):1129-1133.
40. Holmes S, Dickerson J. The quality of life: design and evaluation of a self-assessment instrument for use with cancer patients. *Int J Nurs Stud*. 1987;24(1):15-24.
41. Joyce CR, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur J Clin Pharmacol*. Aug 14 1975;8(6):415-420.
42. Muza SR, Silverman MT, Gilmore GC, Hellerstein HK, Kelsen SG. Comparison of scales used to quantitate the sense of effort to breathe in patients with chronic obstructive pulmonary disease. *Am Rev Respir Dis*. 1990;141(4 Pt 1):909-913.
43. Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain*. Dec 1975;1(4):379-384.
44. Paul-Dauphin A, Guillemin F, Virion JM, Briancon S. Bias and precision in visual analogue scales: a randomized controlled trial. *Am J Epidemiol*. Nov 15 1999;150(10):1117-1127.
45. Shields BJ, Cohen DM, Harbeck-Weber C, Powers JD, Smith GA. Pediatric pain measurement using a visual analogue scale: a comparison of two teaching methods. *Clin Pediatr (Phila)*. Apr 2003;42(3):227-234.
46. Thomas T, Robinson C, Champion D, McKell M, Pell M. Prediction and assessment of the severity of post-operative pain and of satisfaction with management. *Pain*. 1998;75(2-3):177-185.
47. van Laerhoven H, van der Zaag-Loonen HJ, Derckx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr*. Jun 2004;93(6):830-835.
48. Wilson RC, Jones PW. A comparison of the visual analogue scale and modified Borg scale for the measurement of dyspnoea during exercise. *Clin Sci (Lond)*. 1989;76(3):277-282.
49. Youngblut JM, Casper GR. Single-item indicators in nursing research. *Res Nurs Health*. 1993;16(6):459-465.
50. Bowling A. *Research methods in health: investigating health and health services*. Buckingham; Philadelphia: Open University Press; 1998.
51. Murray CJL. *Summary measures of population health: concepts, ethics, measurement and application*. Geneva: World Health Organization; 2002.
52. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain*. Feb 1994;56(2):217-226.
53. Grunberg SM, Groshen S, Steingass S, Zaretsky S, Meyerowitz B. Comparison of conditional quality of life terminology and visual analogue scale measurements. *Qual Life Res*. Feb 1996;5(1):65-72.
54. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med*. Jan 2001;33(1):47-48.
55. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for

- chronic and experimental pain. *Pain*. Sep 1983;17(1):45-56.
56. Maxwell C. Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. *Br J Clin Pharmacol*. Jul 1978;6(1):15-24.
  57. Kaplan GA, Camacho T. Perceived health and mortality: a nine-year follow-up of the human population laboratory cohort. *Am J Epidemiol*. 1983;117(3):292-304.
  58. Miilunpalo S, Vuori I, Oja P, Pasanen M, Urponen H. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *J Clin Epidemiol*. May 1997;50(5):517-528.
  59. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. Mar 1997;38(1):21-37.
  60. Akerstedt T, Knutsson A, Westerholm P, Theorell T, Alfredsson L, Kecklund G. Sleep disturbances, work stress and work hours: a cross-sectional study. *J Psychosom Res*. Sep 2002;53(3):741-748.
  61. Arnetz BB. Staff perception of the impact of health care transformation on quality of care. *International Journal for Quality in Health Care*. 1999;11(4):345-351.
  62. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological Bulletin*. 1955;52:281-302.
  63. Riva G, Teruzzi T, Anolli L. The use of the internet in psychological research: comparison of online and offline questionnaires. *Cyberpsychol Behav*. Feb 2003;6(1):73-80.
  64. Andersson G, Kaldo-Sandstrom V, Strom L, Stromgren T. Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. *J Psychosom Res*. Sep 2003;55(3):259-262.
  65. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis*. Aug 1978;37(4):378-381.
  66. Gorin AA, Stone AA. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In: Baum A, ed. *Handbook of health psychology*. Mahwah, N.J.; London: Lawrence Erlbaum Associates; 2001:405-413.

**Table 1. Background Characteristics of the Respondents**

<b>Variable</b>	<b>N*</b>	<b>% of total</b>
<i>Age (years)</i>		
=30	53	8
31-45	254	40
=46	326	52
<i>Gender</i>		
Male	176	28
Female	457	72
<i>Marital status</i>		
Married /co-inhabiting/live apart	512	81
Single	121	19
<i>Education</i>		
Primary/High school	228	36
Academic degree (BS/BA)	348	55
Higher academic degree (PhD and MA/MS)	57	9
<i>Income per annum US\$</i>		
<30,000	173	27
30,000-45,000	327	52
>45,000	133	21

\*N=633

**Table 2. Questions, Topics and Indices Covered by the Questionnaire**

VAS & Likert items	VAS answer alternatives	Likert answer alternatives
SRH: How would you rate your general state of health?	Poor – Very good	Poor, Quite poor, Neither good or bad, Good, Very good
Health status right now, during last month and last year.	Very poor – Very good	Very poor, Poor, Neither good or bad, Good, Very good
Overall sleep quality?	Very poor – Very good	Very poor, Poor, Neither good or bad, Good, Very good
Quality of sleep right now, last half-year and last year.	Very poor – Very good	See index KSQ.
Energy level right now, last month and last year.	No energy – Full of energy	See index mental energy.
Intensity and frequency of work-related exhaustion	Not at all – Maximum and Never – Most often	See index Work-related exhaustion
Likert-based indices		Answer alternatives
<b>Sleep (KSQ - Karolinska Sleep Questionnaire)</b> <i>Have you experienced any of the below inconveniences during the past time (half-year)?</i> 1.Difficulties to fall asleep, 2.Difficulties waking up, 3.Repeated awakenings with difficulties to fall asleep again, 4.Vigorous snoring (according those around you), 5.Too little sleep (at least one hour less than my sleeping need), 6.Nightmares, 7.Feeling of not being thoroughly rested when you wake up, 8.Too early awakening, 9.Disturbed/anxious sleep,10.Feeling of exhaustion when you wake up, 11.Tired/sleepy at work or during leisure-time, 12.Irritated/tired eyes, 13.Unintentional sleeping episodes (nodding off) at work, 14.Unintentional sleeping episodes (nodding off) during leisure-time, 15.Need to struggle against the sleep to stay awake.		Never, Seldom (some occasions per year), Sometimes (some occasions per month), Mostly (many times per week), Always (every day).
<b>Mental energy (index)</b> <i>Have you experienced any of the below inconveniences during the past month?</i> 1.Feelings of restlessness, 2.Irritability, 3.Worry/anxiety, 4.Difficulty concentrating, 5.Feeling low/moodiness.		No – Yes, some time – Yes, many times – Yes, daily.
<b>Work-related exhaustion (index)</b> <i>How often does the following occur?</i> 1.Emotionally drained after work, 2.Worn out after work, 3.Tired when I think of work.		Never, A few times per year, A few times per month, A few times per week, Daily.

**Table 3. Factor Analyses and Reliability Analyses of the Likert Based Indices**

<b>Index</b>	<b>Cronbach's a</b>	<b>Factor loadings</b>
Disturbed awakening	.81	.42-.81
Disturbed sleep	.76	.47-.86
Sleepiness	.78	.67-.87
Mental energy	.84	.64-.79
Work-related exhaustion	.87	.76-.92

**Table 4. Intraclass Correlations and Differences in Absolute Levels between VAS and Likert Items/Indices**

VAS - Likert item or index	Intraclass correlation r	Marginal Homogeneity test p	Median VAS-Likert %
Self-rated health (VAS) – Self-rated health (Likert)	.90***	ns <sup>a</sup>	70 – 75 <sup>b</sup>
Health right now (VAS) – Health right now (Likert)	.91***	<.05 <sup>a</sup>	68 – 75 <sup>b</sup>
Health last month (VAS) – Health last month (Likert)	.91***	<.05 <sup>a</sup>	65 – 75 <sup>b</sup>
Health last year (VAS) – Health last year (Likert)	.91***	ns <sup>a</sup>	64 – 75 <sup>b</sup>
Overall sleep (VAS) – Overall sleep (Likert)	.94***	ns <sup>a</sup>	65 – 75 <sup>b</sup>
Sleep last half-year (VAS) – Disturbed awakening index (Likert)	.66***	<.001	61 - 50
Sleep last half-year (VAS) – Disturbed sleep index (Likert)	.79***	<.01	61 - 60
Sleep last half-year (VAS) – Sleepiness index (Likert)	.44***	<.001	61 - 75
Energy last month (VAS) – Mental energy index (Likert)	.75***	ns	56 - 60
Work-related exhaustion intensity (VAS) – Work-related exhaustion index (Likert)	.80***	<.001	61 - 42
Work-related exhaustion frequency (VAS) – Work-related exhaustion index (Likert)	.83***	<.001	61 - 42

\*\*\* p<.001 (2-tailed)

<sup>a</sup> For this Marginal homogeneity test the VAS was recoded into categories ranging from 1-5.

<sup>b</sup> The median score for the Likert scale as well as for the recoded VAS is 4. The median figures depicted in the table show the results from when the Likert scales are converted into percentages.