Arbitrary Self-Assessment Scale of Stress: Analysis and Discussion of the limited Relevance

Fauquet-Alekhine, Ph.[1,2,3]; Rouillac, L.[2]
[1] Nuclear Power Plant of Chinon, Head Management Department, BP80 – 37420 Avoine, France
[2] Lab. for Research in Sc. of Energy, Montagret, France
[3] Dept. of Social Psychology, London School of Economics and Political Science, Houghton St., WC2A 2AE, London, UK
email: philippe.fauquet-alekhine@edf.fr [1], larsen.sciences@yahoo.fr [2], p.fauquet-alekhine@lse.ac.uk[3]

Fauquet-Alekhine, Ph; Rouillac, L. (2015) Arbitrary Self-Assessment Scale of Stress: Analysis and Discussion of the limited Relevance, *Stress Self-assessment & questionnaires: choice, application, limits*, 15-20
http//hayka-kultura.org/larsen.html

Abstract

In the aim of assessing subjects' perceived stress, some professionals are seduced by easy arbitrary rating scales which are elaborated for a one-off need out of any scientific and rigorous approach, sometimes called "numerical stress scale". The present research provides an insight into what sort of results may be expected from this kind of assessment. Research articles in peer reviewed journals providing cases of assessment of short-term stress through an arbitrary scale were analyzed and compared with studies applying scientifically validated questionnaires for self-assessment of stress. This objectified the poor reliability of the former compared to the latter and led to identification of weaknesses and improvement suggestions. A calibrating item for arbitrary scales was proposed for future validation.

1. Introduction

Self-assessment is an approach widely applied by professionals in Human Science in order to investigate the perception of subjects involved in a given context. Among these professionals, researchers are used to undertaking such self-assessments on the basis of previous scientific studies providing questionnaires which were tested and validated with several hundreds of subjects according to a strict protocol. Furthermore, these professionals do not usually use questionnaire in language other than that of the original one if the translated questionnaire has not been validated in these new languages. This contributes to validate the influence of another language as well as the influence of another culture associated with the language.

However, some professionals (even among researchers) are seduced by easy arbitrary rating scales which are elaborated for a one-off need out of any scientific and rigorous approach, sometimes called "numerical stress scale". One of the advantages of this type of scale is the following quick analysis: just one number per subject and per condition whereas questionnaires imply several numbers to calculate a score per subject and per condition. For example, Orsila et al. (2008) used what they called a "traditional questionnaire" described in their paper as follows (p.278): "A single survey item was used to assess perceived mental stress, which was elicited on a visual analog scale (from 0—very little stress to 10—very high stress)".

What results may be expected from this kind of assessment? Which level of confidence may be given to this kind of assessment? What kind of bias may affect the results?

The present short paper aims at giving elements of answer to these questions in the case of the assessment of short term stress by analyzing results available in the literature obtained with application of an arbitrary scale and compared with those provided by a scientifically validated method.

2. Material and methods

The stress which we addressed in this analysis was "short term mental stress", not "long term mental stress" linked with periodic stress factors exposure (for example refer to the work of Maslova et al. (2002) who studied the effect of chronic stress on arterial blood pressure, or studies of Schubert et al. (2009) who compared both kinds of stress). Most of the time, short term mental stress (sometimes referred to as "acute stress") occurs whilst dealing with an intense cognitive demand during a short time where "intense" is here taken in a broad range of sense.

It was first necessary to gather data. A bibliographic research aimed at identifying research articles in peer reviewed journals providing cases of stress assessment exclusively through an arbitrary scale for self-assessment of the stress state. For each article, we gathered characteristics of use of this type of scale and analyzed the results and conclusions obtained by the authors.

Arbitrary scales being not scientifically validated, we then presented characteristics of validated questionnaires for self-assessment of stress in order to emphasize the poor quality and reliability of the

former compared to the latter. To illustrate the reliability of validated questionnaires, we summarized a few examples of application of these questionnaires and compared the quality of the results obtained to what studies applying arbitrary scales produced.

3. Results

3.1 Arbitrary scale for self-assessment of stress in the literature

Papers presenting data of self-stress assessment using arbitrary scale are not numerous in the literature, perhaps due to the poor quality of information provided by this method. We only found four articles in peer reviewed journals (however there was one associated to the proceedings of a conference) in which authors used this kind of scale to argue their results regarding stress assessment. In each study, self-assessment of stress through arbitrary scale was compared with one or several physiological techniques of stress assessment but not with another subjective technique; even Goette et al. (2015), who applied the STAI-T questionnaire (Spielberger, 1983, 1989; see below §3.2) to evaluate subjects' trait anxiety and to analyze possible correlations with subjects' state did not apply the STAI-S aiming at evaluation the subjects' anxiety state whereas this questionnaire is scientifically validated. The authors did not explain why they preferred using an arbitrary scale.

The characteristics of stress conditions investigated in the selected papers are summarized in table 1. For all conditions of table 1, subjects were healthy adults, male or female.

Table 1: Characteristics of selected studies

Source	Stress conditions	*N*
Langewitz et al. (1987)	work context vs home context	about 30
Orsila et al. (2008)	occupational work	about 30
Geeraerts et al. (2010)	difficult clinical situation management on anesthesia simulator	about 30
Goette et al. (2015)	interview and Mathematics	about 200

For all conditions, stress was self-assessed using an arbitrary scale of the type described in section "Introduction".

Table 2 provides for each of these studies information about stress assessment techniques that were used in parallel to this arbitrary scale.

For each of these selected studies, the authors' conclusions regarding stress assessment were as follows.

Table 2: Stress assessment methods used in selected studies

Source	Stress assessment method
Langewitz et al. (1987)	blood pressure heart rate respiratory frequency
Orsila et al. (2008)	heart rate heart rate variability
Geeraerts et al. (2010)	salivary amylase
Goette et al. (2015)	salivary cortisol heart rate

Langewitz et al. (1987) compared two conditions of stress exposure for the whole sample: home and work. They concluded that, for the overall, all quantities assessing stress (objective as well as subjective) significantly showed a higher level of stress at work and that ANOVA revealed a significant effect of perceived stress on HR.

Orsila et al. (2008) was the only team presenting calculated correlation between the perceived stress through arbitrary scale and other physiological quantities used as objective assessment of stress (table 3 of their paper). They presented results for quantities generally used to evaluate stress:

- For *HR*, r=-0.41 (p=0.08) with 83% of data for which $p \geq 0.1$.
- For *LF/HF* of heart rate variability, r=-0.3 (p=0.2) with 83% of data for which $p \geq 0.1$.

and they presented results for untypical quantities used to evaluate stress:

- For the Baseline width of the *RR* interval histogram, r=0.73 (p=0.01) with 33% of data for which $p \geq 0.1$.
- For the root mean square of differences of successive *RR* intervals, r=-0.60 (p=0.04).

The authors concluded that "no single parameter seems to correlate with perceived stress" (p.282).

Geeraerts et al. (2010) compared the stress state of subjects just before and just after a given stressful situation for the whole sample: they concluded that they had a very good discrimination with a significant higher level of stress perceived after experiencing the situation than before. They did not mention any correlation between subjective and objectives quantities. However, for 11% of their

data, the perceived stress after experiencing the stressful situation was surprisingly lower than before.

Goette et al. (2015) separated participants into two groups (about 100 subjects each); one group experienced several stressful situations while the other, the control group, did not. They found that "the stress group exhibited higher subjective stress ratings as well as higher cortisol and heart rate level than the control group throughout the session. For subjective stress ratings, there was no significant difference between control and stress groups at the beginning nor at the end of the session. For subsequent measurements, there was a difference between the two groups" (p.118). Regarding physiological measurements, "stress participants had higher salivary cortisol levels than control participants in all measurements" except at the beginning: "there was no difference in salivary cortisol levels between groups"; this was complemented with the fact that "the heart rate differed between the control and stress groups from the minute after the start of the measurement" (p. 118-119).

The arbitrary scale helped the authors to discriminate the stress state of two samples (N>100 each) but they could not obtain discrimination in all the stress conditions studied using this scale: cortisol and HR were more discriminating. The authors did not present any analysis of correlation between subjective and physiological assessment of stress.

3.2 Characteristics of validated questionnaires for self-assessment of stress

There are few scientifically validated questionnaires available for self-assessment of stress. We provide hereinafter a list of such questionnaires with the scientific articles that make reference for each and we summarize the results obtained after the validation process.

ALES, Appraisal of Life Events Scale was elaborated by Ferguson et al. (1999).
The number of items for the questionnaire is 16, selected with reference to the four primary evaluation forms described by Folkman and Lazarus (1985).
The number of participants involved in the validation process is N=260 for exploratory analysis and N=344 for confirmative analysis, giving the total amount N=604.
It has good discriminative sensitivity and good theoretical validity. The internal validity of each factor is satisfactory (α = 0.94 to 0.99), as well as the reproduction through one month test-retest (r = 0.77 to 0.90, p < 0.01) as well as three-month test-retest (r = 0.49 to 0.59, p < 0.01). Regarding external validity, ALES factors are correlated significantly and relevantly with various criteria jointly evaluated (Ferguson et al., 1999).

STAI, State-Trait Anxiety Inventory was elaborated by Spielberger (Spielberger, 1983, 1989; Spielberger & Reheiser, 1994).
The number of items for the questionnaire is 40.
The number of participants involved in the validation process is N>5000.
The STAI intends to assess subjects' conscious awareness at two extremes of anxiety affect, labeled state anxiety (A-State), and trait anxiety (A-Trait), respectively. The original Form X of the STAI was revised resulting in Form Y, a more popular version with improved psychometric properties.
Internal consistency coefficients are satisfactory, ranging from 0.86 to 0.95; two-month test-retest reliability coefficients have ranged from 0.65 to 0.75 (Spielberger et al., 1983).

PDI, Peritraumatic Distress Inventory (Brunet et al., 2001) was elaborated in order to obtain a quantitative measure of the level of distress experienced during and immediately after a traumatic event.
The number of items for the questionnaire is 13.
The number of participants involved in the validation process is N=1003.
It includes the subject's feelings regarding physiological parameters (sweating, shaking, pounding heart). The problem for this questionnaire is that it is linked with the diagnosis of posttraumatic stress disorder (PTSD) which requires that subjects had high levels of distress during or after the traumatic event. This may be a drawback when subjects are submitted to a low level of stress: the questionnaire may not discriminate subjects' stress state (Fauquet-Alekhine et al., 2014).

JSS, Job Stress Survey elaborated by Spielberger (Spielberger, 1983, 1989; Spielberger & Reheiser, 1994) is devoted to perceived stress in professional context.
The number of items for the questionnaire is 60 for the original version and 40 for the shortened version keeping the more stable items (Spielberger & Reheiser, 1994).
The number of participants involved in the validation process is N=1781.
JSS permits to obtain three scores (intensity, frequency, overall) within three domains: job pressure, lack of support, the combination of which referring to occupational stress.
The internal consistency was good while repeatedly scored around 0.80 for the alpha coefficient, and test-retest coefficient was reported at 0.48 to 0.75.
Regarding stress at work, we may also notice the JCQ, Job Content Questionnaire (Karasek et al.,

1998) made of 21 items distributed in 5 domains and tested with more than N=15000 subjects.

For information regarding long term stress, we may notice the followings.

PSS, Perceived Stress Scale was elaborated by Cohen et al. (1983). The number of items for the questionnaire is 14.
The number of participants involved in the validation process is *N*>2300 (40% male) selected as representative of the North-American population (gender, age, income, ethnic, profession).
It has good discriminative sensitivity, a good theoretical validity, with positive correlations with other objective or perceived stress scales. Internal validity is satisfactory (Cohen & Williamson, 1998) and external validity shows that PSS correlates significantly and positively with various indicators of disease, among which the Psychosomatic Index of Derogatis et al. (1976). The questions in this scale ask subjects about their feelings and thoughts during the last month and thus are rather adapted for long term mental stress assessment.

EMS, Experience Sampling Method (Myin-Germeys et al., 2009; Vaessen et al, 2015) allows for in-the-moment assessment of subjective appraisal of a situation and the stress response reflected in current subjective distress or increases in negative affect and symptomatology.

WWBQ, Work and Well-Being Questionnaire (Kilminster et al., 2007; Bridger et al., 2011) was elaborated to measure occupational stressors in the navy.

4. Discussion

It is remarkable that, among all the selected arbitrary scale-based studies, the only one mentioning correlation calculation between subjective and objective quantities provided a very low level of correlation for almost all quantities and especially for all usual quantities used to evaluate stress (*HR*, *LF/HF*). We may assume here that other studies did not dare mentioning any correlation calculation because of their poor quality as reported in §3.1.

No study regarding consistency of arbitrary scale-based questionnaire has been found in the literature probably because of their poor reliability. This poor reliability is obviously mainly due to the high possible disparity which may appear between subjects submitted to the same conditions of stress. This assertion is well illustrated by Geeraerts et al.' data the analysis of which is summarized in §3.1; data are given for each subject in their paper: the variance of perceived stress data related to "before" stress situation is about 30% of the full scale and 50% when related to "after"; this is 63% and 70% of the mean values respectively.

On the contrary, using scientifically validated questionnaire gives reliable data that leads to consistent results. Figure 1 gathers data from two different studies providing data describing subjects in different stressful conditions. This data are from Stedmon et al. (2015) and Berger et al. (2016). Stedmon et al. provided 4 points from the analysis of *N*=38 subjects. Berger et al. provided 4 points from the analysis of *N*=80 subjects. Stedmon et al. analyzed subjects trying to conceal knowledge from interrogators leading them to experience raised levels of stress. Berger et al. recruited subjects to undergo either a social stress situation or a non-stressful control situation. Both studies provided data comparing scores obtained with STAI-S vs HR for groups of subjects in different stressful conditions. When plotted together (Fig. 1), the correlation coefficient is $r(N=8) = 0.95$, $p < 0.0001$. Compared with data of Orsila et al. (2008) summarized in §3.1, Here a clearly higher quality was found for correlation between self-assessment of stress and a physiological quantity illustrating a stress state. This level of quality is not so surprising when taking into account all that was done to validate the STAI questionnaire and the fact that stress is assessed through 40 items instead of a single one.

According to our analysis, the main problem of arbitrary scales for self-assessment of stress is associated with its lack of calibration: even if two subjects feel the same level of stress in a given situation, they may score it at different levels over the scale. For a sample of subjects, this leads to a large variance of the scores which contributes to the poor quality of results obtained with this type of questionnaire. Spielberger (Spielberger, 1983, 1989; Spielberger & Reheiser, 1994) coped with this problem by providing a calibrating item: for the JSS (presented in §3.2), the intensity index is scored over a 1-9 scale, the medium value 5 being calibrated with the first item referring to a subject submitted to unpleasant tasks. This calibrating item leads subjects to score the following items higher than 5 if they are considered more stressful than the calibrating item by the subjects; it is scored less than 5 if the opposite should apply.

For the arbitrary scale of self-assessment, a calibrating level must at least comply with the following properties:

- It must relate to a medium level of stress to make a higher or a lower assessment possible.
- It must concern as much healthy adult subjects as possible, regardless gender, age, profession, social status.

- It must be comparable to other stress factors in as many contexts as possible.

It is clear that this is not easy to fulfill all the criteria at the same time.

The calibrating item referring to the medium level of the scale could be: "you have a letter to send in emergency by post office but you do not have any stamp for it; you are in center town and you know you just have time to go in the last shop which will sell to you the stamp so that your letter might be put in the letter box today to have the stamped date of the day".

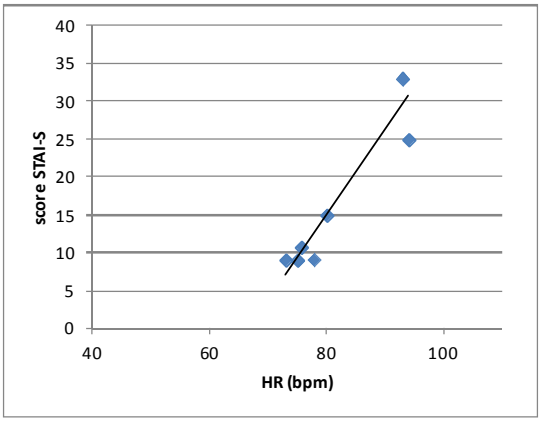

Fig. 1. STAI-S vs HR for groups of subjects in different stressful condition (from Stedmon et al., 2015 and Berger et al., 2016).

When analyzed in the light of aforementioned criteria, this suggested item to calibrate the scale leads to few comments regarding the aforementioned properties that contribute to its limitation:

- Relate to a medium level of stress: anyway, even though many people are concerned by the suggested item, there will be a bias due to personality (as for JSS) and due to culture. This last point might present a reduced limitation as usually tests are undertaken for a sample of subjects concerned by a single culture, making data comparable from one subject to another.
- Concern as much healthy adult subjects as possible: again, even though many people are concerned by the suggested calibrating item, there will be a bias due to personality.
- Be comparable to other stress factors in as many contexts as possible: regarding this point, only a study testing an arbitrary scale applying the suggested calibrating item may objectify a possible bias.

5. Conclusion

Analysis of research articles in peer reviewed journals providing cases of stress assessment exclusively through an arbitrary scale for self-assessment of the state of stress showed the poor reliability of this type of scale. Correlations with physiological parameters were rarely provided by authors, and when they were, the coefficients and significance were low.

A short (and not exhaustive) review of validated questionnaires for self-assessment of stress was undertaken, showing how carefully these questionnaires were elaborated and studied, involving several hundreds or thousands of subjects. An example of correlation between one questionnaire of this type, the State-Trait Anxiety Inventory (STAI), and physiological parameters was given. Two independent studies were used to provide data and calculation showed high correlation, thus illustrating the benefits of scientifically validated questionnaires.

The main problem of arbitrary scales for self-assessment of stress is associated with its lack of calibration: even if two subjects feel the same level of stress in a given situation, they may score it at different levels over the scale. For a sample of subjects, this leads to large variance of the score which contributes to the poor quality of this type of questionnaire. A calibrating item referring to the medium level of the scale is therefore suggested: "you have a letter to send in emergency by post office but you do not have any stamp for it; you are in center town and you know you just have time to go in the last shop which will sell to you the stamp so that your letter might be put in the letter box today to have the stamped date of the day". However validation remains to be carried out. This may be a future research project.

Symbols & Units

Symbol	*Quantity*	*Units (SI)*
N	Sample size	unit
α	Cronbach coefficient	none
p	Probability	none
r	Correlation coefficient	none
HR	Heart Rate	bpm
LF/HF	Low and high frequency ratio of heart rate variability	none
RR	rhythm-to-rhythm (RR) interval	ms

References

Berger, J., Heinrichs, M., von Dawans, B., Way, B. M., & Chen, F. S. (2016). Cortisol modulates men's affiliative responses to acute social stress. *Psychoneuroendocrinology*, *63*, 1-9.

Bridger, R.S., Brasher, K., & Kilminster, S.A.D. (2011).. Job stressors in naval personnel serving on ships and in personnel serving ashore over a twelve month period. *Applied Ergonomics*, 42, 710-718.

Brunet, A., Weiss, D., Metzler, T., Best, S., Neylan, Th., Rogers, C., Fagan, J., & Marmar, Ch. (2001). The Peritraumatic Distress Inventory: A Proposed Measure of PTSD Criterion A2. *Am J Psychiatry*, 158(9), 1480-1485.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396.

Cohen, S.; Williamson, G.M. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan et S. Oskamp (eds), p. 31-67. *The social psychology of health*. London : Sage publications.

Derogatis, L., Rickels, K. et Rock, A. (1976). The SLC-90 and the MMPI: a step in the validation of a new self-report scale, *British Journal of Psychiatry*, 128, 280-289.

Fauquet-Alekhine, Ph.; Geeraerts, Th.; Rouillac, L. (2014) Characterization of anesthetists' behavior during simulation training: performance versus stress achieving medical tasks with or without physical effort, *Psychology and Social Behavior Research*, 2(2) 20-28

Ferguson, E., Matthews, G., Cox, T. (1999). The appraisal of life events (ALE) scale : reliability and validity, *The British Psychological Society*, 4, 97-116.

Folkman, S. et Lazarus, R. S. (1985). If it changes it mustbe a process : study of emotion and coping during three stages of a college examination, *Journal of Personality and Social Psychology*, 48, 150-170.

Geeraerts, T., Roulleau, P., Cheisson, G., Aidan, K., Leguen, M., Duranteau, J., & Benhamou, D. (2010). Stress induced by simulation-based training in anesthesiologist residents. *Critical Care*, 14, 1-2.

Goette, L., Bendahan, S., Thoresen, J., Hollis, F., & Sandi, C. (2015). Stress pulls us apart: Anxiety leads to differences in competitive confidence under stress. *Psychoneuroendocrinology*, 54, 115-123.

Karasek, R.; Brisson, Ch.; Amick, B. (1998) The Job Content Questionnaire (JCQ): An Instrument for Internationally ComparativeAssessments of Psychosocial Job Characteristics, *Journal of Occupational Health Psychology*, 3(4), 322-355

Kilminster, S., & Bridger, R.S. (2007). Modified work and Well-Being Questionnaire for Cohort *Study of Stress in the Naval Service: Psychometric Validation*. Unpublished MoD Report.

Langewitz, W., Rüddel, H., & Von Eiff, A. W. (1987). Influence of perceived level of stress upon ambulatory blood pressure, heart rate, and respiratory frequency. *Journal of clinical hypertension*, 3(4), 743-748.

Maslova, L.N. , Bulygina,V.V. , & Markel, A.L. (2002). Chronic stress during prepubertal development: immediate and long-lasting effects on arterial blood pressure and anxiety-related behavior. *Psychoneuroendocrinology*, 27, 549-561.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Eperience sampling research in psychopathology: opening the black box of daily life. *Psychological Medicine*, 39, 1533-1547.

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., Nygård, C.H. (2008). Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics*, 14(3), 275-283.

Schubert, C., Lambertz, M., Nelesen, R.A., Bardwell, W., Choi, J.B., & Dimsdale, J.E. (2009). Effects of stress on heart rate complexity—A comparison between short-term and chronic stress. *Biological Psychology*, 80, 325-332.

Spielberger, C.D. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologist Press Inc., Palo Alto, California.

Spielberger, C.D. (1989). *State-Trait Anxiety Inventory: Bibliography* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., & Reheiser, E. C. (1994). The job stress survey: Measuring gender differences in occupational stress. *Journal of Social Behavior and Personality*, 9(2), 199-218.

Stedmon, A. W., Eachus, P., Baillie, L., Tallis, H., Donkor, R., Edlin-White, R., & Bracewell, R. (2015). Scalable interrogation: Eliciting human pheromone responses to deception in a security interview setting. *Applied ergonomics*, *47*, 26-33.

Vaessen, Th.; Van Nierop, M.; Reininghaus, U.; Myin- Germeys, G. (2015) Stress Assessment Experience Sampling: Convergent Validity and Clinical Relevance, *Stress Self-assessment & questionnaires: choice, application, limits,* 21-35