

The Development of a Spanish Language Version of the Worn Out Scale of the General Well-Being Questionnaire (GWBQ)

Tom Cox¹, Amparo Oliver², Eusebio Rial-González³,
José Manuel Tomás², Amanda Griffiths¹, and Louise Thompson¹

¹Institute of Work, Health & Organizations, University of Nottingham (UK)

²Universitat de València

³Agencia Europea para la Seguridad y la Salud en el Trabajo, Bilbao (Spain)

The paper describes the development of a short Spanish-language version of the General Well-Being Questionnaire (GWBQ; Cox & Gotts, 1987), based on the 12 items of its Worn Out scale. Research has shown the English-version Worn Out scale to be sensitive to aspects of the design and management of work. This study aimed to test its cross-cultural consistency in a Spanish-language workplace context. The data were collected from a sample of 229 workers in Valencia (Spain). Confirmatory Factor Analyses showed the factorial validity, reliability, and concurrent validity of the new Spanish version to be adequate. The sensitivity of the new measure to safety behavior and the reporting of accidents was also assessed and shown to be good. The new questionnaire extends the usefulness of the parent questionnaire to occupational health psychology research in the Spanish language by offering a short assessment tool appropriate for workplace studies. *Keywords: well-being, Spanish-language, occupational health, workplace research*

El artículo describe el desarrollo de la versión corta en español del General Well-Being Questionnaire (GWBQ; Cox & Gotts, 1987; en español el "Cuestionario de Bienestar General"), basada en los 12 ítems de su subescala de "Agotamiento". La investigación ha mostrado que la versión en inglés de la escala de Agotamiento es sensible a ciertos aspectos del diseño y gestión del trabajo. El propósito de este estudio fue comprobar la consistencia transcultural en un contexto de trabajo de habla española. Los datos se recogieron de una muestra de 229 trabajadores en Valencia (España). Análisis factoriales confirmatorios demostraron la adecuación de la validez factorial, la fiabilidad, y la validez concurrente de la nueva versión española. También se evaluó la sensibilidad de la nueva medida hacia la conducta de seguridad y distintas medidas de siniestralidad, comprobando asimismo su adecuación. El cuestionario nuevo amplía la utilidad del cuestionario original al ofrecerse como herramienta para los investigadores en psicología de la salud ocupacional en lengua española, ofreciendo un instrumento corto de evaluación apropiado para el lugar de trabajo.

Palabras clave: bienestar, español, salud ocupacional, investigación en psicología del trabajo

Health has been defined as a changeable state along a continuum from complete well-being to death (Rogers, 1960; World Health Organization, 1946). This broad definition implies that health cannot be equated with mere absence of obvious disease, and is not restricted to the sound physical condition of the body, but encompasses also psychological and social factors. In this context, *well-being* relates to individuals' experience of their health status.

An important watershed in the health continuum is represented by the point where disease, disability, or injury become obvious to the person, and are usually represented by objectively verifiable and clinically significant signs and symptoms. Some of these will be diagnostic of a particular condition; others will more generally reflect the impact of that condition. Some will inevitably represent the effects of stress experienced in relation to being ill. The watershed marks one divide between positive and negative well-being. *Positive well-being* is conceptualized as more than a lack of disease, injury, and disability. The measurement of positive well-being is one of the future challenges for occupational health psychology. At present, most measures offer some form of count of signs and symptoms of ill-health: relative well-being is defined by a low or zero count.

The zone between complete physical, psychological, and social well-being on the one hand, and obvious disease, disability, or injury, on the other, has been termed *suboptimal health* (Rogers, 1960). It has been suggested that suboptimal well-being may be represented as an experiential pool of signs and symptoms of general malaise, each—on its own—of no particular clinical significance and certainly not diagnostic of any particular condition. Such signs and symptoms may or may not be precursors to disease, injury, or disability depending on the operation of a wide range of health risk and salutogenic factors. At any time, different groups of signs and symptoms within the experiential pool will imperfectly predict particular ill-health outcomes. As a condition develops, the predictive group will refine itself, attract new signs and symptoms, and the prediction itself may strengthen. As a condition weakens, or the person recovers, then the reverse process will occur.

A person's perception of their own health (well-being) is shaped by three factors: (a) awareness and evaluation of symptoms of general malaise, (b) awareness and evaluation of indicators of positive well-being, and (c) diagnosis (self- or formal) and/or awareness of a particular disease, injury, or disability. The first two factors co-exist, and the perception of general well-being is some sort of balancing or dynamic appraisal process, elements of which are rescaled by the third factor before reaching a final judgment.

A person considered to be normally healthy, by himself or by others, or more particularly, judged not to be ill, will still experience something by way of such signs and symptoms of general malaise, and the possible pool of such experiences will be formally present at the group level. At this level, structural modeling will reveal a pattern and

clusters of signs and symptoms reflective of an underlying *normal experience model*. However, it has been suggested that the normal experience of well-being may both reflect the experience of stress as one mediator of the effects of life and working conditions, and also in turn affect other responses to stress, such as self-reported mood (see Mackay, Cox, Burrows, & Lazzarini, 1978; Cox & Mackay, 1985).

Development of the General Well-Being Questionnaire (GWBQ; Cox & Gotts, 1987)

In the mid 1980s, the Nottingham group began to build a theoretical model of well-being that was deliberately developed from that implicit in the World Health Organization's (1946) definition of health, and drew on the work of Rogers (1960). The group also developed an associated measurement tool based on the self-report of signs and symptoms of general malaise that could measure suboptimal health as defined earlier.

Initially, a compilation of nonspecific symptoms was produced from existing health questionnaires and from scrutiny of diagnostic texts. These symptoms included reportable aspects of cognitive, emotional, behavioral, and psychological function indicative of general malaise, none of which were clinically significant in themselves. From this compilation, a prototype checklist was designed with each symptom being associated with a 5-point frequency scale (*never* through to *all the time*) that referred to the respondent's experience during the previous 6 months. In a series of exploratory factor analytical studies with British participants, using the psychometric good practice described by Ferguson and Cox (1993), two clusters of symptoms or factors were identified. These factors were derived as orthogonal. The first factor (GWF1) was defined by symptoms relating to tiredness, emotional lability, and cognitive confusion; it was labeled *worn out*. The second factor (GWF2) was defined by symptoms relating to worry and fear, tension and physical signs of anxiety; it was labeled *tense and uptight*. This model of suboptimal well-being was shown to have face validity in that it was acceptable to a conference audience of British general practitioners and medical researchers (see Cox, Thirlaway, Gotts, & Cox, 1983).

This research culminated in the publication in English of the General Well-Being Questionnaire (Cox & Gotts, 1987). In the late 1980s, new data were collected through a series of linked studies in Britain and Australia. These data were reanalyzed, and the model and its associated scales were amended to increase their robustness in relation to this international sample, and also to a diversity of homogeneous samples. A number of symptoms (items) were deleted from the original scales, but no new symptoms were added. The two new international scales were each defined by twelve symptoms but retained their essential nature: worn out and

tense and uptight. The deleted symptoms were among the weaker ones in terms of scale definition and item loadings. The early questionnaire was revised, new norms were computed and an international version was published and has been in use since then (e.g., Cox & Griffiths, 1996, Cox, Griffiths, Barlow, Randall, Thomson & Rial-González, 2000).

Measuring Well-Being at the Workplace

The measurement of well-being in workplace has tended to focus on three different parts of the health continuum: (a) the incidence of death (mortality), (b) of work-related disease, disability, or injury (morbidity), and (c) of general malaise. It appears from many studies in occupational health psychology that the effects of work design and management are more likely to be expressed in terms of changes in well-being than in the incidence of death, or of disease, disability, or injury (see Cox, Griffiths, & Rial-González, 2000). It is therefore important for this discipline to have tools that can reliably and validly measure the well-being of working people. Several measures have been developed to this end, but many of these are open to criticism particularly with regard to their purpose, their length, their cultural specificity, and their psychometric qualities.

In the context of the increasing trend to trans-European research, several of the established measures are of questionable structural validity when used outside their original cultural-linguistic context. For example, factorial studies of the 12-item General Health Questionnaire (Goldberg, 1972) in Spain have not shown a clear factor structure: neither a two-factor structure based on anxiety and depression, nor a single general factor (Oliver, Sancerni, Tomás, & Lis, 1995). Whereas other measures show good structural reliability, some have not demonstrated adequate discriminant validity. Many are simply too long, which is an important disadvantage in workplace studies.

The possibility of cultural-linguistic differences in the experience and report of well-being has been explored in working populations in Taiwan (Ruey-Fa, 1994) and in Singapore (Ho, 1996). For example, Bilingual Taiwanese school teachers (English and Mandarin) completed English and Mandarin versions of the GWBQ. The data clearly showed the emergence of identical two factor models from these data sets with teachers' scores on the two versions being very highly correlated. The Taiwanese (Mandarin) model was indistinguishable from the UK English language version. Despite this extreme test, the cultural-linguistic inter-changeability of the GWBQ should not be taken for granted, and is a matter of empirical test between its English origins and other cultural-linguistic situations.

This paper explores data from a Spanish working population both to satisfy a practical need and to answer a question; first, there is a need in occupational health psychology research in Spain for a *short* tool with which to measure well-being in workplaces, and, second, there is

the real question of whether the adaptation of the GWBQ in order to fulfill that need would reveal cultural-linguistic effects as discussed above.

Logically, it was possible to shorten the existing instrument in two different ways: by reducing the length of both scales (Worn Out and Tense and Uptight), or by selecting one of the two scales. The second strategy was favored, and the Worn Out scale chosen, for two main reasons: First, there are insufficient data on the reliability and validity of shortened versions of the two scales. Second, there is some evidence that in workplace studies, the Worn Out scale shows at least equal reliability and external validity as the Tense and Uptight scale, and greater utility, this is, a more consistent relationship with other, *nonhealth* measures of interest within the workplace (Cox et al. 2000a).

The Worn Out scale of the GWBQ was translated into Spanish and administered to a heterogeneous population of Spanish workers along with a number of other established psychometric instruments. The internal structure of the new Spanish short-version GWBQ was then examined, and its reliability, validity, and utility explored.

Method

Participants

The present study is based on results drawn from a sample of 229 valid returns. Participants' ages ranged from 16 to 64 years ($M = 36.4$, $SD = 11.0$). Out of the sample, 86.5% were men. Their educational level was mainly elementary (61.8%); 11.8% of the sample had attended secondary school, 18.2% had achieved technical qualifications, and 8.2% held university degrees.

With regard to job level, 75.4% of the sample were classed as workers, 19.3% were supervisors, and 5.3% managers. Nearly half the sample (46.6%) had permanent contracts, 26.5% held temporary contracts, and 27% had some form of casual contract. The mean job tenure of participants was 7.8 years ($SD = 8.3$). Participants worked, on average, 8.4 hours overtime per month ($SD = 4.9$).

All the participants were employed in small or medium sized enterprises with no more than 100 workers: the mean for organization size was 22 employees.

Measures

The following instruments were included in the survey questionnaire:

1. A socio-demographic section, including measures of age, gender, educational level, job status, and type of contract.
2. An employment section including measures of length of employment in general, period in the current company, period in the actual job, working hours per week, and overtime per month.

3. A work-related safety section including measures such as number and severity of injuries suffered during the previous 2 years. The safety of workers' behavior was measured using a 5-item, 4-point Likert scale that ranged from 1 (*never safe*) to 4 (*always safe*) ($\alpha = .513$).

4. Three health measures were used in the study:

4.1. The 24-item GWBQ showed an internal consistency of $\alpha = .921$, measured by Cronbach's alpha, including both the Tense and Uptight and Worn Out scales (Cox & Gotts, 1987), both with a 5-point Likert scale (0 = *never*, 1 = *rarely*, 2 = *sometimes*, 3 = *often*, 4 = *all the time*). The Worn Out scale had an α of .858 and the Tense and Uptight scale had an α of .829. This study focuses on the data from the Worn Out scale.

4.2. A 9-item anxiety checklist ($\alpha = .826$) was also administered to participants using a response scale from 1 (*clearly yes*), 2 (*just a few*), 3 (*not sure*) to 4 (*clearly not*). All statements referred to the extent to which respondents felt the adjectives defined their mood when completing the survey. The items' response scale was reversed for purposes of analysis where necessary.

4.3 The 12-item General Health Questionnaire (*GHQ-12*; Goldberg, 1972) showed an overall reliability of $\alpha = .756$, with $\alpha = .470$ for its Anxiety subscale, and $\alpha = .696$ for the Depression subscale. The response scale for these items was a 4-point Likert scale with the following responses: 1 (*not at all*), 2 (*no more than usual*), 3 (*more than usual*), and 4 (*much more often than usual*).

Procedure

This study presents a psychometric analysis of responses to a Spanish language version of the Worn Out scale of the General Well-Being Questionnaire (GWBQ) when administered to a heterogeneous sample of workers from small or medium sized enterprises in the province of Valencia, Spain.

The workers were interviewed and the questionnaires administered during their annual occupational health examinations at the Regional Government's Health and Safety Offices. A team of doctors performed the medical examinations, conducted the interviews, and administered the questionnaires. Each worker completed the questionnaire individually and returned it in a sealed envelope to ensure anonymity. Data were collected by random sampling of the examined population over the course of one year.

The survey was part of a larger project investigating health and safety issues carried out by the Faculty of Psychology and the Faculty of Medicine (University of Valencia), in collaboration with the Government of Valencia.

Statistical Analyses

The analyses focused on the 12 Worn Out items of the GWBQ. Descriptive statistics were calculated and the reliability (internal consistency) of the scale and of the individual items was analyzed. The scale data were then

correlated with a number of external health and safety measures. Correction for attenuation was applied to all the correlations of Worn Out with these measures (Nunnally & Bernstein, 1994). All these analyses were performed using the SPSS 12 statistical package. An alpha level of .05 was used for all statistical tests.

Confirmatory factor analysis. Confirmatory factor analysis (CFA)-EQS 5.1 (Bentler, 1995)-was used to test the factorial validity of the Worn Out scale because an a priori model of well-being existed (see above).

A critical issue in any confirmatory factor analysis is the assessment of model fit that gives the technique its confirmatory function. A number of different goodness-of-fit indices exist. They can be grouped into different families according to a shared rationale. As no single measure of fit can be relied on exclusively, researchers are advised to use a number of different indices (Hoyle, 1995; Tanaka, 1993). The indices selected here were: the chi-square statistic, Tucker-Lewis index (TLI), comparative fit index (CFI), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), and the root mean square error of approximation (RMSEA).

The most widely used measure of fit is the chi-square test, although it suffers from several shortcomings: among others, it is based on highly restrictive assumptions, it depends on sample size, and models are an approximation to reality rather than an exact representation (e.g., Bentler & Bonnet, 1980). Because of these problems, a number of other indices have to be consulted as well.

The GFI is an absolute fit index developed by Jöreskog and Sörbom (1984) with a corresponding adjusted version, the AGFI, developed to incorporate a penalty function for the addition of free parameters in the model. These indices are measures of the relative amount of variance and covariance in the data matrix that are accounted for by the model. The GFI is analogous to R-square and performs better than any other absolute index (Hoyle & Panter, 1995; Marsh, Balla, & McDonald, 1988).

RMSEA is the square root of the ratio of the re-scaled noncentrality index to the model's degrees of freedom (Loehlin, 1998). A value of the RMSEA above 0.1 would cast doubt on the model fit, about 0.08 to 0.05 would indicate a reasonable error of approximation, and a value of about 0.05 or less would indicate a close fit (Browne & Cudeck, 1993).

TLI is a Type-2 incremental fit index and compares the lack of fit of a target model to the lack of fit of a baseline model, usually an independence model. The CFI is a Type-3 incremental fit index that also assesses the adequacy of a target model in relation to a baseline model, but defining a population fit index parameter and then using estimators of this parameter to define sample fit indices. Both TLI and CFI have values between 0 and 1, with 1 indicating perfect fit. A value of .9 has usually been considered as a minimum for model acceptance. In this study, those indices

with the best behavior according to the literature have been selected (Gerbing & Anderson, 1993; Hoyle & Panter, 1995; Hu & Bentler, 1995; Marsh, Balla, & Hau, 1996; Oliver & Tomás, 1995).

The items in the Worn Out scale refer to a sample of symptoms and signs that can be absent or present with varying frequency. This makes nonnormality and categorization problems likely. These sorts of problems affect structural equation models in general, but confirmatory factor analyses in particular.

The standard method of estimation in structural equation models is maximum likelihood (ML), which assumes that the variables in the model are continuous and that they have multivariate normal distribution. Univariate normal distribution is a necessary but not sufficient condition to achieve multivariate normality. Theoretically, two problems occur with ML estimators if variables depart from normality (West, Finch, & Curran, 1995): The chi-square tests reject too many true models, and tests of the parameter estimates are biased, yielding too many significant results. Moreover, crude categorization of continuous variables, as is the case when frequency symptom is measured with a Likert scale, leads to biased chi-square tests, parameter estimates, and standard errors (e.g., Bollen & Barb, 1981). Simulation studies on the effects of nonnormality have reached several conclusions (Browne, 1984; Hu, Bentler, & Kano, 1992; West et al., 1995): (a) ML estimation produces positively biased chi-square when the data become increasingly nonnormal, (b) nonnormality produces modest underestimation of fit indices such as TLI and CFI, and (c) standard errors of parameter estimates present moderate to strong negative bias. However, according to the available Monte Carlo studies (Babakus, Ferguson & Jöreskog, 1987; Boomsma, 1983; Muthen & Kaplan, 1985, 1992), the categorization of continuous variables does not produce a large effect on the chi-square test when variables do not depart from normality markedly, although, as the distributions become increasingly and differentially skewed, the chi-square values are positively biased. The Monte Carlo evidence also suggests that there is a slight underestimation of the parameter estimates when distributions are not multivariate normal, and that this increases when there are few response categories (less than four), while skewness increases. In summary, departures from normality, together with an extreme categorization of continuous variables, can lead to important estimation problems and an incorrect evaluation of model fit.

There are a number of remedies available, based either on alternative methods of estimation or on some type of re-expression of variables. Among the first type of solution, distribution free methods (Browne, 1984), a scaled chi-square statistic and robust standard errors (Satorra, 1990), bootstrapping techniques, and the continuous/categorical variables methodology (CVM) developed by Muthen (1984) are available within the structural equation modeling

framework. The second type of solution includes either item parceling (Bagozzi & Heatherton, 1994) or the transformation of the variables to approximate normality (Daniel & Wood, 1980; Dunn, Everitt, & Pickles, 1993). All these alternatives were considered and applied in this study where necessary and possible for a better estimation of the parameters in the model.

Results

Descriptive Statistics and Reliability

Table 1 presents the descriptive statistics for the 12 Worn Out items. In general, item means are low. The highest mean (1.532) belonged to item number 7 ("Have things got on your nerves and worn you out?"). Items 2 and 11 also had relatively high means. Item number 9 ("Has your face become flushed?") had the lowest mean, 0.605. In terms of percentages, 56.8% of the sample replied "never" to this item, and 24.8% "rarely"; a response pattern consistent with the item's low mean and positive skewness. Table 1 also includes skewness and kurtosis statistics for each item. This table shows that, as usual for this type of symptom scale, the variables do not approximate normality. Most of the items presented positive skewness of varying magnitude. Only items number 4 and 7 presented a slight negative skewness.

Corrected item-total scale correlations are also presented in Table 1. These correlations give an indication of the internal consistency of the scale, and therefore, their reliability in measuring feelings of being worn out. Item-scale correlations ranged from .35 to .65. Most of the items presented correlations with the scale above .50, and only two presented correlations around .45. The item least consistent with the scale was item number 4 ("Have you become easily annoyed or irritated?"), which is also the most skewed item. Cronbach's α for the Worn Out scale was .86.

Factorial Validity

Multivariate kurtosis was high, with a Mardia's coefficient of 13.76, and normalized estimate of 4.71. The size of the Mardia coefficient indicates that the hypothesis of multivariate normality cannot be sustained. Therefore, using standard normal theory estimators with these data could produce estimation problems. On the other hand, the Likert scale seems adequate. The number of categories used, five, seems safe for the estimation techniques available. The main statistical problem was the severe nonnormality of the data. In this situation, application of the available remedies for nonnormality was required prior to the use of the standard maximum likelihood estimation procedure. Free distribution methods and CVM estimators were not suitable due to their sample size requirements (at least 500 observations). Scaled chi-square test and robust statistics

Table 1

Descriptive Statistics: Mean, Standard Deviation, Skewness, Kurtosis, Item-total Correlation and Factor Loading

Symptoms	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Item-total Correlation	Factor Loading
1. bored	0.937	0.935	0.809	0.200	.499	.531
2. annoyed or irritated	1.510	0.991	0.029	-0.704	.650	.677
3. clear throat	0.826	1.006	1.025	0.126	.353	.455
4. thinking got mixed	1.189	0.864	-0.082	-1.141	.535	.668
5. impulse	0.610	0.814	0.892	-0.556	.433	.511
6. forgotten things	1.195	0.860	0.265	-0.640	.440	.553
7. worn out	1.532	0.974	-0.127	-0.540	.558	.622
8. tired	1.042	0.907	0.538	-0.255	.607	.700
9. face flushed	0.605	0.808	0.973	-0.223	.537	.599
10. sleeping problems	1.068	1.003	0.547	-0.671	.590	.635
11. feelings easily hurt	1.447	1.071	0.031	-0.882	.616	.688
12. hard to make up mind	1.279	0.987	0.231	-0.798	.543	.665

can correct the overall fit of the model and the standard errors of parameter estimates, but not the estimation of the parameters themselves. On the other hand, item parceling was not considered adequate in this case because the particular information of the items is important and the aim of the analysis is to establish the factorial validity of the scale. Therefore, after careful evaluation of these procedures, it was decided to transform the original variables in order to approximate normality. A square-root transformation was chosen because it is the least intrusive transformation available for positively skewed variables such as those in the Worn Out scale. The transformation produced, as expected, a dramatic change in the multivariate kurtosis (West et al., 1995): Mardia's coefficient was then 6.11, with a normalized estimate of 2.09.

The development of the Nottingham model of well-being and the results obtained with the English-language version of the questionnaire suggest a one-factor solution for the items that make up the Worn Out scale. In order to test this hypothesis, a maximum likelihood estimated confirmatory factor analysis was carried out on the transformed data. The confirmatory factor analysis was specified with a single factor explaining every individual item. According to the goodness-of-fit indices, the one factor solution can be safely maintained, $\chi^2(54) = 107.668$, $p < .001$, TLI = .90, CFI = .912, GFI = .897, AGFI = .851, and RMSEA = .08. The CFI and TLI were both around .9. GFI also evaluated the model as adequate with a value close to .9 as well. The AGFI was slightly lower. The AGFI penalizes complex models, and it is quite useful in model comparison. The model proposed here is, however, extremely parsimonious because it comprises 12 observed variables in one single factor. RMSEA, also a parsimony index, was .08, indicating a reasonable error of approximation. In summary, the one factor solution represents a good model fit.

Factor loadings offer information on the reliability of the items to measure the Worn Out Factor, and an indication of the analytic fit of the model. All factor loadings were statistically significant ($p < .001$). Standardized factor loadings are shown in Table 1. All items were highly related to the Worn Out factor. The lowest value was for item 3 (.455) but all remaining loadings were above .5. In general, items were excellent indicators of the latent variable.

Concurrent Validity: Health Criteria

In order to analyze the concurrent validity of the scale, Worn Out scores were correlated with other scales that purport to measure well-being. In particular, the Worn Out scores were correlated with the total score on the General Health Questionnaire (Goldberg, 1972), as well as with its Anxiety and Depression subscales. They were also correlated with an anxiety checklist. The correlations with corrections for attenuation are shown in Table 2. The correlations with these health criteria were adequate, indicating acceptable concurrent validity for the new Spanish Worn Out scale.

Utility and Validity: Safety Criteria

The Worn Out scale has been used in different contexts either as a dependent variable related to the assessment of work characteristics, or as a predictor of other health-related outcomes. The English-language version of the Worn Out scale has shown great utility in dealing with health-related (and work-related) issues (Cox et al. 2000a). In this study, the scale has shown statistically significant correlations with a number of safety indicators. Table 2 shows the correlations with these criteria, as well as their corrections for attenuation. The Worn Out scale data were significantly related to the number of accidents experienced in a 2-year period, scores

Table 2

Criterion Validity of Worn Out: Correlation Coefficients (r), Their Statistic Probabilities (p) and Correction for Attenuation (ρ)

		Worn out		
		<i>r</i>	<i>p</i>	ρ
Utility criteria	Accidents	.167	.023	.180
	Hazards	.167	.029	.180
	Safe behavior	-.189	.013	-.285
	Overtime	.181	.013	.195
	Age	-.157	.032	-.169
Health criteria	GHQ	.430	< .001	.534
	GHQ anxiety	.444	< .001	.699
	GHQ depression	.311	< .001	.402
	Anxiety checklist	.473	< .001	.562

on a hazards awareness checklist, on a safeness of behavior measure, and to the extent of overtime working (hours). The directions of the correlations were as hypothesized, showing that the Worn Out scale was sensitive to changes in the safety context. However, as might have been anticipated, these correlations were not as high as those with other well-being criteria.

Discussion

The study reported here was successful in allowing the development of a Spanish language version of the Worn Out scale of the GWBQ for use as a short (12 item) instrument to measure well-being. In doing so, it has demonstrated two important points: First, translating this scale from its original English cultural-linguistic context to a Spanish one did not substantively alter its structure or its behavior in relation both to well-being and safety measures. Second, the scale has some degree of validity in measuring self-reported well-being, and of utility in the understanding of safety in the workplace. This is important given that the management of psychosocial hazards has been recently included in Spanish health and safety legislation with the transposition of the European Union's Framework Directive.

The new Spanish scale was shown to have adequate internal consistency and a robust factor structure, and also to have adequate concurrent validity to support its use as a measure of well-being. Interestingly, the Worn Out scale was shown to correlate more strongly with the Anxiety scale of the GHQ than with its Depression scale: This finding requires further exploration with both English and Spanish samples, but suggests that simply equating the Worn Out construct with either anxiety or depression is unsatisfactory. As a construct, worn out has much in common with the notion of exhaustion or fatigue broadly defined. Evidence reported elsewhere (e.g., Schaufeli, Maslach, & Marek, 1993) has demonstrated correlations of about .4 between scores

on the Worn Out scale and those on the emotional exhaustion scale of the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981). The Worn Out scale has been designed to measure non-specific symptoms of general malaise defining sub-optimal health. However, the GHQ is more concerned with the measurement of minor psychiatric morbidity focused on two specific states: anxiety and depression. The two measures must logically overlap, but represent both different levels of granularity in measurement and different *zones* in the health continuum. Within the framework of the model proposed earlier, the closer the person is to the watershed of diagnosis awareness of the anxiety or depression states, the stronger the correlation with Worn Out scores will be.

In this light, the correlations between the Worn Out scale of the GWBQ and the GHQ are respectable, and support differential usage of the two instruments. The same conclusion can be reached, but for different reasons, in relation to the Emotional Exhaustion scale of the MBI: whereas the Worn Out scale is context-free, the Emotional Exhaustion scale of the MBI is anchored to a particular type of work context. They measure different things, but must, and do, demonstrate a reasonable correlation.

The Worn Out scale assesses feelings of being exhausted, cognitively confused, and emotional labile over the previous six months. Feelings of being worn out appear to be related to the self-reported safeness of workers' behavior, to the number of accidents that they report having had, and to their awareness of hazards to their safety and health. The pattern of correlation, while implying no particular causal relationships, describes a situation encompassing workers who are worn out, behaving relatively unsafely, unaware of the hazards of work, working extra hours, and reporting a relatively high level of accidents within the sample. This pattern has some face validity within the context of European safety management experience.

Workplace studies in occupational health and safety generally require measuring instruments that are as short as possible while remaining consistent with the purpose of the study. Long questionnaires seem to pose two related challenges

to researchers: low completion and return rates, and, as a result, data of questionable quality. Both detract substantively from the overall quality and reliability of any study. A reliable and valid 12-item scale of well-being has obvious attractions for use in workplace studies. The Spanish version of the Worn Out scale offers such an instrument. Future research should extend the use of the new short questionnaire to other Spanish samples and explore further the association between Worn Out scores and health and safety related measures, particularly in relation to the assessment of the characteristics of work design and management.

References

- Babakus, E., Ferguson, C.E., & Jöreskog, K.G. (1987). The sensitivity of confirmatory factor maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 37, 72-141.
- Bagozzi, R.P., & Heatherton, T.F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling*, 1, 35-67.
- Bentler, P.M. (1995). *EQS for Macintosh user's guide*. Encino, CA: Multivariate Software.
- Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A., & Barb, K.H. (1981). Pearson's *r* and coarsely categorized measures. *American Sociological Review*, 46, 232-239.
- Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality*. Unpublished doctoral dissertation, University of Groningen, Groningen (Germany).
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Cox, T., & Gotts, G. (1987). *The General Well-Being Questionnaire Manual*. Nottingham, UK: University of Nottingham, Department of Psychology.
- Cox, T., & Griffiths, A.J. (1996). The assessment of psychosocial hazards at work. In M.J. Schabracq, J.A.M. Winnubst, & C.L. Cooper (Eds.), *Handbook of work and health psychology*. Chichester, UK: Wiley.
- Cox, T., Griffiths, A.J., Barlow, C.A., Randall, R.J., Thomson, L.E., & Rial-González, E. (2000). *Organisational interventions for work stress*. Sudbury, UK: HSE Books.
- Cox, T., Griffiths, A.J., & Rial-González, E. (2000). *Research on work-related stress*. Luxembourg: Office for Official Publications of the European Communities.
- Cox, T., & Mackay, C.J. (1985). The measurement of self-reported stress and arousal. *British Journal of Psychology*, 76, 183-186.
- Cox, T., Thirlaway, M., Gotts, G., & Cox, S. (1983). The nature and assessment of general well-being. *Journal of Psychosomatic Research*, 27, 353-359.
- Crown, S., & Crisp, A.H. (1966). A short clinical diagnostic self rating scale for psychoneurotic patients. The Middlesex Hospital Questionnaire (MHQ). *British Journal of Psychiatry*, 112, 917-923.
- Daniel, C., & Wood, F.S. (1980). *Fitting equations to data*. New York: Wiley.
- Dunn, G., Everitt, B., & Pickles, A. (1993). *Modelling covariances and latent variables using EQS*. London: Chapman and Hall.
- Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A user's guide. *International Journal of Selection and Assessment*, 1, 84-94.
- Gerbing, D.W., & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.
- Goldberg, D.P. (1972). *The detection of psychiatric illness by questionnaire* [Maudsley Monograph no 21]. London: Oxford University Press.
- Ho, J. (1996). *School organizational health and teacher stress in Singapore*. Unpublished doctoral dissertation, University of Nottingham, UK.
- Hoyle, R.H. (1995). *Structural equation modeling: Concepts, issues and applications*. Newbury Park, CA: Sage.
- Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 158-176). Thousand Oaks, CA: Sage.
- Hu, L.T., & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.
- Hu, L.T., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.
- Jöreskog, K.G., & Sörbom, D. (1984). *LISREL VI user's guide* (3rd ed.). Mooresville, IN: Scientific Software.
- Loehlin, J.C. (1998). *Latent variable models*. New Jersey: Erlbaum.
- Mackay, C.J., Cox, T., Burrows, G. & Lazzarini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social and Clinical Psychology*, 17, 283-284.
- Marsh, H.W., Balla, J.R., & Hau, K.T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315-353). Mahwah, NJ: Erlbaum.
- Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 102, 391-410.
- Maslach C., & Jackson S.E. (1981). *Maslach Burnout Inventory*. Palo Alto, CA: Consulting Psychologist Press.

- Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.
- Muthen, B., & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.
- Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- Oliver, A., Sancerni, M.D., Tomás, J.M. & Lis, R. (1995). Métodos de estimación y tamaños muestrales en análisis factorial confirmatorio: Implicaciones en la validez factorial del GHQ. *Psicológica*, *16*, 101-114.
- Oliver, A., & Tomás, J.M. (1995). Índices de ajuste absolutos e incrementales: comportamiento en análisis factorial confirmatorio con muestras pequeñas *Psicológica*, *16*, 1, pp.49-64.
- Rogers, E.H. (1960). *The ecology of health*. New York: MacMillan.
- Ruey-Fa, L. (1994). *Organisational healthiness, stress, and well-being*. Unpublished doctoral dissertation, University of Nottingham, UK.
- Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity*, *24*, 367-386.
- Schaufeli, W. & Maslach, C., & Marek, T. (Eds.) (1993). *Professional burnout: Recent developments in theory and research*. New York: Hemisphere.
- Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.
- West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with non-normal variables. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*, (pp.56-75). Thousand Oaks, CA: Sage.
- Wilson, J., & Corlett, N. (1995). *The evaluation of human work: A practical ergonomics methodology*. London: Taylor and Francis.
- World Health Organization (1946). *World Health Organization: Basic Documents* (26th ed.). Geneva: World Health Organization.

Received, March, 8, 2005

Review received, September, 14, 2005

Accepted: October, 21, 2005