Routledge
Taylor & Francis Group

Psychosocial risk assessment in organizations: Concurrent validity of the brief version of the Management Standards Indicator Tool

Jonathan Houdmont[a]*, Raymond Randall[b], Robert Kerr[c] and Ken Addley[d]

[a]*School of Medicine, University of Nottingham, UK;* [b]*School of Business and Economics, Loughborough University, Loughborough, UK;* [c]*Department of Management and Leadership, University of Ulster, UK;* [d]*Northern Ireland Civil Service Occupational Health Service, Belfast, UK*

The Management Standards Indicator Tool (MSIT) is a 35-item self-report measure of the psychosocial work environment designed to assist organizations with psychosocial risk assessment. It is also used in work environment research. Edwards and Webster presented a 25-item version of the MSIT based on the deletion of items having a factor loading of < .65. Stress theory and research suggest that psychosocial hazard exposures may result in harm to the health of workers. Thus, using data collected from three UK organizations ($N = 20,406$) we compared the concurrent validity of the brief and full versions of the MSIT by exploring the strength of association between each version of the instrument and a measure of psychological wellbeing (GHQ-12 and Maslach Burnout Inventory). Analyses revealed that the brief instrument offered similar but not always equal validity to that of the full version. The results indicate that use of the brief instrument, which would be less disruptive for employees, would not elevate the risk of false negative or false positive findings in risk assessment.

Keywords: Management Standards Indicator Tool; psychosocial work environment; work-related stress; validity

Introduction

The UK Health and Safety Executive's Management Standards Indicator Tool (MSIT) is a freely available 35-item measure of exposure to seven dimensions of the psychosocial work environment that, if not properly managed, can lead to harm to employees. These dimensions include job demands, job control, managerial support, peer support, relationships, role and change. Initially designed to assist organizations in meeting their legal duty in relation to psychosocial risk assessment (Cousins et al., 2004; Mackay, Cousins, Kelly, Lee, & McCaig, 2004), the instrument's popularity as a work environment research tool has mushroomed in the UK and elsewhere. Studies have focused on the establishment of benchmark scores (Houdmont, Kerr, & Addley, 2012; Houdmont, Kerr, & Randall, 2012; Kinman & Court, 2010; Kumar & Madhu, 2012), the instrument's psychometric properties (Edwards & Webster, 2012; Edwards, Webster, Van Laar, &

---

*Corresponding author. Email: jonathan.houdmont@nottingham.ac.uk

Easton, 2008; Magnavita, 2012; Marcatto et al., 2011; Rondinone et al., 2012; Toderi et al., 2013) and policy applications (Iavicoli, Natali, Rondinone, Castaldi, & Persechino, 2010). Consistent with transactional stress theory that conceptualizes psychosocial hazard exposures as potential contributory factors in health impairment (Cox & Griffiths, 2010), research has also explored associations between psychosocial hazard exposures measured using the MSIT and a variety of health indices (Bartram, Yadegarfar, & Baldwin, 2009; Bevan, Houdmont, & Menear, 2010; Guidi, Bagnara, & Fichera, 2012; Kasi & Haslam, 2013; Kerr, McHugh, & McCrory, 2009; Magnavita, 2012; Marcatto, D'Errico, Di Blas, & Ferrante, 2011; Ravalier, McVicar, & Munn-Giddings, 2013; Toderi et al., 2013). To date, this latter category of studies has explored the MSIT largely in relation to psychological health through the use of many different measurement instruments including the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), the General Wellbeing Questionnaire (Cox, Thirlaway, Gotts, & Cox, 1983), the Job-related Wellbeing Scale (Warr, 1990), the Maslach Burnout Inventory (MBI; Maslach, Jackson, & Leiter, 1996) and the GHQ-12 (Goldberg & Williams, 1988). The GHQ-12 is the only instrument to have been used in more than one study. In line with theoretical expectations, the results of these studies have, in general, indicated that responses on the MSIT scales tend to exhibit acceptable correlations with psychological health. Such findings strengthen the case for using the MSIT to assess the risk to health posed by employees' exposure to psychosocial hazards.

In a recent issue of this journal Edwards and Webster (2012) tested the factor structures of the 35-item version of the MSIT and of a subset of 25-items each with a factor loading of $\geq$ .65. The results showed that both the full and brief versions of the tool offered a good fit to the data. However, there remains an absence of evidence concerning the concurrent validity of the brief 25-item scale relative to that of the full 35-item scale. In this paper we present results from the analysis of three large data sets (combined $N = 20,406$). We used these data sets to test whether the two versions of the questionnaire have equivalent concurrent validity in relation to two different and widely-used measures of psychological wellbeing: the GHQ-12 and the emotional exhaustion scale of the MBI.

There is a strong imperative for brief measurement instruments in occupational health psychology research and organizational psychosocial risk management activities. From the employer perspective, a key benefit of a brief measurement instrument can be found in the savings to employee downtime that it affords. The average time to complete the full 35-item Management Standards Indicator Tool is publicized by the HSE as 13 minutes. On this basis the brief 25-item version would represent a time saving of 3.7 minutes per respondent (assuming all items take approximately equal time to complete). This could lead to considerable savings at the organizational level. For example, for the sample on which the current study is based, application of the brief measure in preference to the full measure would have saved a total of 1263 hours (approximately 34 working weeks). Brief measures are also advantageous in that they have the potential to decrease the risk of participants' data representing a response set, and typically generate a response rate superior to that achieved by lengthy measures. Systematic reviews of response rates achieved in health-focused studies have consistently demonstrated an inverse relationship between questionnaire length and response rate (Edwards et al., 2002; Edwards, Roberts, Sandercock, & Frost, 2004). To summarize, should the current study find that the brief version of the MSIT has equivalent validity to the full version, evidence will have been found to support its use in research and practice.

Method

Participants and procedure

Sample 1. The first sample comprised 17,124 UK civil servants (51% response rate) who contributed data in 2005. Details of the data collection protocol and sample characteristics are presented in Houdmont, Kerr, and Addley (2012). Ethical approval was granted by the Northern Ireland Civil Service (NICS) Workplace Health Committee, as part of the Stormont Study that is tracking a large cohort of employees through their career with the NICS and beyond.

Sample 2. The second sample comprised 1741 police officers drawn from a UK territorial police force (23% response rate). Data collection took place in 2009. Details of the data collection protocol and sample characteristics are presented in Houdmont, Kerr, and Randall (2012). The study was approved by the University of Ulster Research Ethics Committee.

Sample 3. The third sample comprised 2026 police officers from a different UK territorial police force from which Sample 2 was drawn (25% response rate). Data collection took place in 2012. Details of the data collection protocol and sample characteristics are reported in Houdmont (2012). Approval for the study was granted by the Police Federation Joint Branch Board.

There were some missing data in all three samples. Listwise deletion of cases was used in both the correlation and regression analyses (see Table 1 for samples sizes used in correlation and regression analysis).

Measures

Psychosocial hazard exposure. The UK Health and Safety Executive Management Standards Indicator Tool (HSE, n.d.) was used with all three participant samples. The full version includes 35 items and the brief version, in which only items with a factor loading $\geq .65$ are retained, contains 25 items. Responses to most items are given on a five-point scale: 1 (*never*), 2 (*seldom*), 3 (*sometimes*), 4 (*often*) and 5 (*always*). The remaining items involve responses given on a five-point scale of 1 (*strongly disagree*), 2 (*disagree*), 3 (*neutral*), 4 (*agree*) and 5 (*strongly agree*). Items are worded so that low scores are indicative of high (and potentially harmful) exposures. Differences between the full and brief versions of the MSIT are located within four of the seven subscales. These subscales are Demands (full version = eight items and brief version = four items), Control (six and four), Relationships (four and two) and Role (five and three). The remaining three subscales – Managerial Support, Peer Support and Change – are identical in both versions. Exemplar items from each of the seven scales include: "I have unachievable deadlines" (demands), "I have a say in my own work speed" (control), "I am given supportive feedback on the work I do" (managerial support), "If work gets difficult my colleagues will help me" (peer support), "I am subject to bullying at work" (relationships), "I am clear what my duties and responsibilities are" (role) and "Staff are always consulted about change at work" (change).

Mean item scores for the brief scales were calculated for use in the analysis. Data from the MSIT were, in general, normally distributed. In all samples, the relationships scale showed some negative skew, indicating that relatively few participants reported problems with this aspect of their work. This is a common finding (Edwards & Webster, 2012). The large sample sizes meant that transforming the data was not necessary.

Psychological wellbeing. In order to test for concurrent validity, two widely used measures of psychological wellbeing were also applied to the same samples. Samples 1 and 2 completed the 12-item General Health Questionnaire (GHQ-12; Goldberg & Williams, 1988). Items are scored on a four-point scale so that higher scores are indicative of higher levels of psychological distress. A sample item is "[I have] Been able to concentrate on whatever you are doing [over the past few weeks]" 1 (*better than usual*), 2 (*same as usual*), 3 (*less than usual*) and 4 (*much less than usual*). The GHQ scoring method (0-0-1-1) was used to score the data, as advocated by the test author (Goldberg & Williams, 1988). This scoring method was designed to reduce measurement errors that might be introduced by a participant's tendency to endorse extreme responses (or to over-use scale mid-points) and to provide a score that is not influenced by intensity of symptoms (ibid). This scoring method has also been found to be more reliable than the alternative Likert scoring method (Hankins, 2008). Therefore, we used the GHQ scoring method to minimize the standard error of measurement. This allowed us to carry out a more accurate analysis of its correlations with MSIT scales, which were not specifically designed to measure the intensity of stressors. Points were summed to a global score ranging from 0–12, which was used as a continuous variable in the analysis. Reliability was high in both Samples 1 and 2 ($\alpha > = .92$).

Sample 3 was administered the Emotional Exhaustion (EE) subscale of the Maslach Burnout Inventory (MBI). The MBI is available in three forms; this study used the Human Services version (Maslach et al., 1996), which is designed for use with employees in human services jobs. The EE subscale consists of nine items ($\alpha = .90$), each of which is scored on a seven-point response scale for frequency of occurrence (0–6), in which 0 indicates that the statement never applies, through to 6 which indicates that it applies every day. An example item is "I feel emotionally drained from my work". A summed score out of a maximum score of 63 was calculated for each participant.

Analysis

We restricted our analysis to the four scales that differed between versions of the MSIT. We analysed Pearson's correlations between the four scales on each version of the MSIT and the measures of psychological wellbeing. Negative correlations were expected, as reported exposure to poor working conditions should be associated with poor wellbeing as indicated by high GHQ and MBI scores. The equivalence of each pair of correlations was then compared (brief vs. full) using a series of paired *r* tests carried out using Steiger's (1980) formula for comparing non-independent correlations. To test the equivalence of the validity of the scales when both are used in the same study, two linear regression analyses were carried out for each sample. In one regression analysis we used the full scales as predictor variables, and in the other we used the brief scales as predictors (in both regressions the predictor variables were entered in a single step). The

Table 1. Descriptive data and intercorrelations between brief scales of the Management Standards Indicator Tool and measures of wellbeing (GHQ and MBI). Correlations between the *full* MSIT scales and psychological wellbeing are shown in brackets.

	Sample (1–3)	Mean (SD)	1	2	3	4	5
1. Demands	1	3.65 (0.84)	(.80)				
	2	3.17 (0.83)	(.81)				
	3	3.13 (0.79)	(.79)				
2. Control	1	3.25 (0.80)	.31**	(.79)			
	2	2.97 (0.80)	.41**	(.82)			
	3	2.93 (0.80)	.38**	(.81)			
3. Relationships	1	4.50 (0.81)	.35**	.30**	(.77)		
	2	4.34 (0.81)	.29**	.26**	(.74)		
	3	4.31 (0.86)	.29**	.26**	(.71)		
4. Role	1	3.96 (0.83)	.26**	.36**	.27**	(.77)	
	2	3.88 (0.77)	.39**	.37**	.29**	(.80)	
	3	3.74 (0.82)	.29**	.33**	.22**	(.82)	
5. GHQ-12[a]	1	2.51 (3.52)	−.31** (−.31**)	−.27** (−.28**)	−.31** (−.36**)	−.28** (−.32**)	(.92)
GHQ-12[a]	2	4.17 (4.10)	−.42** (−.43**)	−.35** (−.36**)	−.37** (−.47**)	−.41** (−.42**)	(.94)
EE MBI[a]	3	26.93 (11.78)	−.10** (−.09**)	−.11** (−.11**)	−.08** (−.09**)	−.10** (−.10**)	(.90)
Paired r t-value			0.00	3.90**	12.68**	17.37**	
			1.33	1.29	8.1**	0.77	
			−1.15	0.00	0.74	0.00	

Notes: Total $N = 20,406$. Sample 1, civil servants, 2005; $N = 16,814$. Sample 2, police officers, 2009; $N = 1741$. Sample 3, police officers, 2012; $N = 1851$. GHQ = General Health Questionnaire; MBI = Maslach Burnout Inventory. [a]Two correlations between the GHQ-12 (Samples 1 and 2) and the EE MBI (Sample 3) and the MSIT scales are given. Figures on the diagonal are scale reliabilities for the brief scale.
**$p < .001$.

multiple *R*s for each model (full vs. brief version) were also then compared using Steiger's (1980) formula as described in Tabachnick and Fidell (2001, pp. 146–147).

Results

Internal consistency was high for the full version of all four scales across all samples: job demands (eight items: Sample 1, $\alpha = .86$; Sample 2, $\alpha = .87$; Sample 3, $\alpha = .84$), job control (six items: Sample 1, $\alpha = .80$; Sample 2, $\alpha = .84$; Sample 3, $\alpha = .82$), relationships (four items: Sample 1, $\alpha = .78$; Sample 2, $\alpha = .76$; Sample 3, $\alpha = .74$) and role (five items: Sample 1, $\alpha = .82$; Sample 2, $\alpha = .85$; Sample 3, $\alpha = .85$). These scale reliabilities were consistent with those found in other large-scale studies that have used the instrument (Bevan et al., 2010; Edwards et al., 2008; Houdmont, Kerr, & Addley, 2012; Houdmont, Kerr, & Randall, 2012; Kerr et al., 2009; Kinman & Court, 2010).

Table 1 presents correlations between variables and scale reliabilities for each of the four brief scales. It shows that these were marginally lower than those for the corresponding full scales but remained above the widely accepted lower threshold of $\alpha = .7$.

Sample 1

In the sample of civil servants (Sample 1), there were modest negative correlations between all four brief MSIT scales and GHQ scores (see Table 1). These were broadly similar across the four scales (from $r = -.27$, $p < .001$ for Control to $r = -.31$, $p < .001$ for Demands and Relationships). When compared to the correlations between the full version of the MSIT and the GHQ, three significantly larger correlations between the GHQ and full measures of Control, Relationships and Support were found but these were of a small magnitude (see Table 1). Inspection of β values in the two regression analyses (see Table 2) also showed a high degree of similarity between the brief and full versions of the scales. The overall adjusted R^2 for the two regressions indicated that the brief scale explained 18% of the variance in GHQ scores, only 2% less than that accounted for by the full scale. A comparison of multiple *R* values indicated that there was a significant difference in prediction offered by the two regression models ($t = -13.86$, $p < .001$).

Sample 2

In the 2009 sample of police officers (Sample 2) the negative correlations between the four brief MSIT scales and the GHQ were relatively large and highly significant (from $r = -.35$, $p < .001$ for Control to $r = -.42$, $p < .001$ for Demands). Fisher's tests indicated that these were all significantly stronger than those found using the brief MSIT in Sample 1 (*z* ranged from 2.69 to 5.87, $p < = .004$). However, the comparison of the correlations between the GHQ and the full and brief versions of the MSIT revealed only one significant difference: the correlation between the brief MSIT Relationships scale and the GHQ was significantly smaller than it was between the full MSIT scale and the GHQ ($t = 8.1$; $p < .001$). The overall adjusted R^2 for the two regressions indicated that the brief scale explained 30% of the variance in GHQ scores, 3% less than that accounted for by the full scale. Compared to the regression using the full scales, there was a significant

Table 2. Comparison of beta values for full and brief scales of the Management Standards Indicator Tool for all three samples.

	Sample 1: Civil servants ($N = 16,814.$ DV = GHQ-12)		Sample 2: Police officers ($N = 1741.$ DV = GHQ-12)		Sample 3: Police officers ($N = 1851.$ DV = MBI EE)	
	Full	Brief	Full	Brief	Full	Brief
Demands	−.16**	−.18**	−.21**	−.23**	−.03	−.05
Control	−.07**	−.11**	−.08**	−.13**	−.07*	−.07*
Relationships	−.20**	−.18**	−.27**	−.21**	−.04	−.03
Role	−.18**	−.15**	−.20**	−.21**	−.06*	−.05*
R	.45	.43	.57	.55	.14	.14
Adjusted R^2	.20	.18	.33	.30	.02	.02
F	1040.28**	941.06**	212.53**	189.57**	8.65**	9.05**

Notes: DV = dependant variable; GHQ = General Health Questionnaire; MBI = Maslach Burnout Inventory. $*p < .05$; $**p < .001$

drop in the multiple R ($t = −3.21$, $p < .001$) when the brief scales were used in the regression model.

Sample 3

In the most recent sample of police officers (Sample 3), there were relatively small but highly significant negative correlations between all four brief scales and the MBI EE, ranging from $r = −.08$, $p < .001$ for Relationships to $r = −.11$, $p < .001$ for Control (see Table 1). There were no significant differences between the full and brief scales in terms of their correlations with the MBI EE scale. The overall adjusted R^2 for the two regressions indicated that both versions of the scale explained just 2% of the variance in emotional exhaustion, with there being no significant decrease in multiple R associated with use of the brief scales.

Discussion

This study set out to explore the concurrent validity of the brief 25-item MSIT instrument relative to that of the full 35-item tool. The results of our analysis indicate that the brief MSIT scales offer similar but not always equal concurrent validity to that provided by the full MSIT scales. This similarity between the scales was found when using two different and widely-used measures of wellbeing as criterion measures, and when correlations between the MSIT and wellbeing were small, modest and large. The small decreases in concurrent validity between the two instruments may be linked to the reduction in content validity and reliability of the subscales of the brief instrument (although the latter reduction was very small, as was also found by Edwards and Webster, 2012) that inevitably occur when shortening a scale. The concurrent validity of the brief version of the Relationships scale may be particularly susceptible to such effects (see Tables 1 and 2). This could be because items in the full Relationship scale that are omitted from the brief version explain unique portions of the variance in wellbeing (i.e. different aspects of working relationships have independent effects on wellbeing). In addition, the sensitivity of the tests that revealed significant differences between the concurrent validity of the two

versions are affected by sample size and therefore small differences were highly significant in Sample 1. Such statistical differences may not have practical significance when decisions are being made about the need for risk reduction activities.

The portion of the variance in Emotional Exhaustion (Sample 3) accounted for by the full and brief versions of the MSIT was identical. In the other two samples the full version of the instrument accounted for 2–3% more variance in the GHQ-12 than the brief version. It could be argued that the additional variance accounted for by the full version relative to the brief version is noteworthy and might offer a rationale for retention of the full version. This may be especially important for researchers seeking to establish the relative importance of different stressors in determining employee wellbeing in circumstances where a lower response rate can be tolerated. However, in the context of organizational psychosocial risk assessment activities – for which the MSIT was developed – such differences may be of less importance. This is because the objective in this context is to produce an overall group-level snapshot of workers' psychosocial hazard exposures. This can then be used to highlight aspects of the psychosocial work environment that are identified as problematic by the statistical majority of respondents and which might therefore warrant prioritization within a stressor reduction intervention programme. Applied to this purpose, the additional variance accounted for by the full version is unlikely to influence significantly the decision-making process when key stakeholders consider the need for action. Our results indicate that use of the brief scales would not elevate the risk of false negative findings in the risk assessment process for Samples 1 and 2. It was also notable that the use of the brief version did not inflate the small correlations observed in Sample 3. Thus, it appears that there is also a low risk of false positive results associated with use of the brief scale.

The weak correlations between the two versions of the MSIT and emotional exhaustion observed for Sample 3 may also indicate that it did not adequately assess the specific psychosocial hazards that are linked to this criterion variable in this sample. This lends support to the argument that generic psychosocial work environment measures may not always be sensitive to the hazards associated with particular specialized occupations such as policing. Other occupation-specific hazards may need to be added to improve the content validity of the instrument when used in police contexts where emotional exhaustion is a problem. The same might be true for other occupations. Indeed, there is some evidence to suggest that the augmentation of generic psychosocial hazard measures with occupation-specific measures may contribute to the explanation of additional variance in health outcomes. For example, de Croon et al. (2002) found that Karasek's (1979) job demand-control model accounted for 22% of the variance in fatigue among lorry drivers; the addition of job-specific demands to the model resulted in the explanation of an additional 3% of the variance. The extent to which this might be the case where the MSIT is applied remains unclear but the results shown in Table 2 indicate that this possibility should be explored. To this end, the first two authors are currently undertaking a nationwide longitudinal study among police officers that includes various police-specific psychosocial hazard measures alongside the MSIT.

Our analysis revealed very different relationships between the MSIT and wellbeing in three samples. This suggests that it is important to include measures of wellbeing in addition to those of psychosocial hazards in risk assessment activities, as the concurrent validity of the scales can differ significantly across samples. At a practical level, use of the brief scales provides an opportunity to include measures of wellbeing at the same time without making the assessment process excessively time-consuming and disruptive.

However, the impact of the loss of breadth in the brief version of the measure on the quality of information available in the intervention design process will also need to be examined in future research.

One of the strengths of this study is that we were able to draw on large data sets obtained from very different organizational contexts. This indicates that the results may well generalize to other study populations, although this will need to be tested in future studies. The equivalence of the concurrent validity of the two scales also needs to be examined in a large private-sector sample and with other criterion measures (such as absence, turnover and performance). A natural extension of our study would be to conduct a longitudinal assessment of the predictive validity of the two versions of the MSIT. This would help to control for methodological issues such as common method variance and help to rule out the possibility of reverse causality.

Conclusion

In conclusion, this investigation has shown the brief version of the MSIT to have broadly equivalent concurrent validity to the full version. Our findings indicate that this brief version, which is more efficient and less disruptive to participants, can be used in occupational health psychology research and organizational psychosocial risk management activities.

References

Bartram, D., Yadegarfar, G., & Baldwin, D. (2009). Psychosocial working conditions and work-related stressors among UK veterinary surgeons. *Occupational Medicine*, *59*, 334–341.

Bevan, A., Houdmont, J., & Menear, N. (2010). The Management Standards Indicator Tool and the estimation of risk. *Occupational Medicine*, *60*, 525–531.

Cousins, R., Mackay, C., Clarke, S., Kelly, C., Kelly, P., & McCaig, R. (2004). 'Management Standards' and work-related stress in the UK: Practical development. *Work & Stress*, *18*, 113–136.

Cox, T., & Griffiths, A. (2010). Work-related stress: A theoretical perspective. In S. Leka & J. Houdmont (Eds.), *Occupational health psychology* (pp. 31–56). Oxford: Wiley-Blackwell.

Cox, T., Thirlaway, M., Gotts, G., & Cox, S. (1983). The nature and assessment of general wellbeing. *Journal of Psychosomatic Research*, *27*, 353–359.

de Croon, E., Blonk, R., de Zwart, B., Frings-Dresen, M., & Broersen, J. (2002). Job stress, fatigue, and job dissatisfaction in Dutch lorry drivers: Towards an occupation specific module of job demands and control. *Occupational and Environmental Medicine*, *59*, 356–361.

Edwards, J., & Webster, S. (2012). Psychosocial risk assessment: Measurement invariance of the UK Health and Safety Executive's Management Standards Indicator Tool across public and private sector organizations. *Work & Stress*, *26*, 130–142.

Edwards, J., Webster, S., Van Laar, D., & Easton, S. (2008). Psychometric analysis of the UK Health and Safety Executive's Management Standards Work-Related Stress Indicator Tool. *Work & Stress*, *22*, 96–107.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: Systematic review. *British Medical Journal*, *324*, 1183–1193.

Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials: Does questionnaire length matter? *Controlled Clinical Trials*, *25*, 31–52.

Goldberg, D., & Williams, P. A. (1988). *User's guide to the General Health Questionnaire*. Windsor: NFER-Nelson.

Guidi, S., Bagnara, S., & Fichera, G. (2012). The HSE Indicator Tool, psychosocial distress and work ability. *Occupational Medicine*, *62*, 203–209.

Hankins, M. (2008). The reliability of the 12-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health*, 8, 3555.

Health and Safety Executive (n.d.). HSE Management Standards Indicator Tool. Retrieved from www.hse.gov.uk/stress/standards/pdfs/indicatortool.pdf.

Houdmont, J. (2012). *West Midlands police officers' exposure to work-related stressors: Relationships with individual and organisational health*. Research report for West Midlands Police Federation. Retrieved from http://westmidspolfed.com/media/documents/wmpfwholeforcesurvey2012.pdf

Houdmont, J., Kerr, R., & Addley, K. (2012). Psychosocial factors and economic recession: The Stormont Study. *Occupational Medicine*, 62, 98–104.

Houdmont, J., Kerr, R., & Randall, R. (2012). Organisational psychosocial hazard exposures in UK policing: Management Standards Indicator Tool reference values. *Policing: An International Journal of Police Strategies and Management*, 35, 182–197.

Iavicoli, S., Natali, E., Rondinone, B., Castaldi, T., & Persechino, B. (2010). Implementation and validation in the Italian context of the HSE Management Standards: A contribution to provide a practical model for the assessment of work-related stress. *Giornale Italiano Di Medicina Del Lavoro Ed Ergonomia*, 32, 130–133.

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24, 285–309.

Kasi, A., & Haslam, C. (2013). Stress management standards: A warning indicator for employee health. *Occupational Medicine*, 63, 335–340.

Kerr, R., McHugh, M., & McCory, M. (2009). HSE Management Standards and stress-related work outcomes. *Occupational Medicine*, 59, 574–579.

Kinman, G., & Court, S. (2010). Psychosocial hazards in UK universities: Adopting a risk assessment approach. *Higher Education Quarterly*, 64, 413–428.

Kumar, K., & Madhu, G. (2012). An analysis of work related stress factor in selected industries in Kerala, India. *International Journal of Engineering Research and Development*, 1, 31–36.

Mackay, C., Cousins, R., Kelly, P., Lee, S., & McCaig, R. (2004). 'Management Standards' and work-related stress in the UK: Policy background and science. *Work & Stress*, 18, 91–112.

Magnavita, N. (2012). Validation of the Italian version of the HSE Indicator Tool. *Occupational Medicine*, 62, 288–294.

Marcatto, F., D'Errico, G., Di Blas, L., & Ferrante, D. (2011). Assessing work-related stress: An Italian adaptation of the HSE Management Standards Work-Related Stress Indicator Tool. *Giornale Italiano Di Medicina Del Lavoro Ed Ergonomia*, 33, 403–408.

Maslach, C., Jackson, S., & Leiter, M. (1996). *Maslach Burnout Inventory Manual* (3rd ed.). Mountain View, CA: CPP.

Ravalier, J., McVicar, A., & Munn-Giddings, C. (2013). The Management Standards Indicator Tool and evaluation of burnout. *Occupational Medicine*, 63, 145–147.

Rondinone, B., Perschino, B., Castaldi, T., Valenti, A., Ferrante, P., Ronchetti, M., & Iavicoli, S. (2012). Work-related stress risk assessment in Italy: The validation study of Health and Safety Executive Indicator Tool. *Giornale Italiano Di Medicina Del Lavoro Ed Ergonomia*, 34, 392–399.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Toderi, S., Balducci, C., Edwards, J., Sarchielli, G., Broccoli, M., & Mancini, G. (2013). Psychometric properties of the UK and Italian versions of the HSE stress indicator tool. *European Journal of Psychological Assessment*, 29, 72–79.

Warr, P. B. (1990). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology*, 63, 193–210.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67, 361–370.