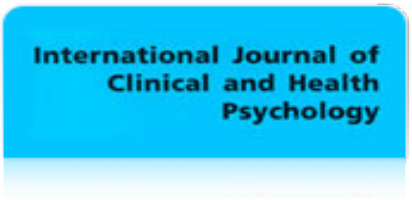

International Journal of Clinical and Health Psychology
ISSN: 1697-2600
jcsierra@ugr.es
Asociación Española de Psicología Conductual
España

Aguayo, Raimundo; Vargas, Cristina; de la Fuente, Emilia I.; Lozano, Luis M.
A meta-analytic reliability generalization study of the Maslach Burnout Inventory
International Journal of Clinical and Health Psychology, vol. 11, núm. 2, 2011, pp. 343-361
Asociación Española de Psicología Conductual
Granada, España

© International Journal of Clinical and Health Psychology

ISSN 1697-2600
2011, Vol. 11, No. 2, pp. 343-361

A meta-analytic reliability generalization study of the Maslach Burnout Inventory[1]

Raimundo Aguayo, Cristina Vargas, Emilia I. de la Fuente[2], and Luis M. Lozano
(*Universidad de Granada, Spain*)

ABSTRACT. The Maslach Burnout Inventory is one of the most widely used measuring instruments for assessing the construct of Burnout and several versions of the test and adaptations to different languages have been developed. This scale measures three dimensions: *Emotional exhaustion*, *Depersonalization* and *Personal accomplishment*. The present meta-analysis is a reliability generalization study examining where the homogeneity of reliability estimates across different empirical studies in each of these dimension. Fifty-one Cronbach's alpha coefficients from 45 empirical studies were analysed, showing an average reliability of .88, .71, and .78, respectively for each dimension. A high level of heterogeneity was found (the I^2 indexes were of 93.7%, 95.5%, and 96.3%). Seven moderator variables were identified to explain the heterogeneity in the *Emotional exhaustion* dimension, and three others in *Depersonalization*. Finally some implications for the empirical studies that use this scale are discussed, concluding that it is inadvisable to continue with the practice of reliability induction when the MBI is administered: the reliability obtained in each application should be included in the empirical studies.

KEYWORDS. Maslach Burnout Inventory. Meta-analysis. Reliability generalization. Cronbach´s alpha. Instrumental study.

RESUMEN. El Maslach Burnout Inventory es uno de los instrumentos de medida más usado para evaluar el constructo de Burnout del que se han construido distintas versiones y adaptaciones a otros idiomas. Este instrumento mide tres dimensiones: *Agotamiento emocional*, *Despersonalización* y *Realización personal*. El presente meta-análisis es un estudio

[1]This work has been funded with the project SEJ2006-13009 of the Ministry of Education and Science and the project P07HUM-02529 of excellence of the Junta de Andalucía.
[2]Correspondence: Facultad de Psicología. Universidad de Granada. Campus de Cartuja, s/n. 18071 Granada (España). E-mail: edfuente@ugr.es

de generalización de la fiabilidad que examina si existe homogeneidad en las estimaciones de dicho parámetro, a través diferentes estudios empíricos, en cada una de estas dimensiones. Se analizaron 51 coeficientes alfa a partir de 45 investigaciones, mostrando una fiabilidad media de 0,88, 0,71 y de 0,78, respectivamente, para las dimensiones citadas. Se obtuvo una elevada heterogeneidad (los índices I^2 fueron 93,7%, 95,5% y 96,3%). Se obtuvieron siete variables moderadoras que explicaban la heterogeneidad en la dimensión *Agotamiento emocional* y tres en *Despersonalización*. Finalmente, se discuten las implicaciones para los estudios empíricos cuando se usa este instrumento, concluyendo que no es aconsejable continuar con la práctica de la inducción de la fiabilidad cuando el MBI es aplicado: la fiabilidad obtenida en cada muestra se debe incluir en los estudios empíricos.

PALABRAS CLAVE. Maslach Burnout Inventory. Meta-análisis. Generalización de la fiabilidad. Coeficiente alfa. Estudio instrumental.

The term Burnout was coined by Freudenberger (1974), the conceptual definition proposed by Maslach and Jackson (1981) being the one most used in the scientific community. These authors defined the Burnout construct as an inappropriate response to chronic work stress that is characterised by Emotional exhaustion (EE), Depersonalization (D) and low Personal accomplishment (PA). The Burnout syndrome manifests itself in those who work in the helping professions; its development brings with it a deterioration in physical and mental health and it has negative consequences in the personal and work spheres.

An appropriate detection of Burnout would allow interventions that minimised its adverse consequences; for this reason, numerous instruments have been devised to measure it, the most widely used being the Maslach Burnout Inventory (MBI) (Maslach and Jackson, 1981; Seisdedos, 1997), that has been used in different cultural contexts and professions (Bernaldo de Quirós-Aragón and Labrador-Encinas, 2007; Jenaro-Río, Flores-Robaina, and González-Gil, 2007; Topa-Cantisano and Morales-Domínguez, 2007).

The first version of the MBI constituted 22 items that evaluate the three dimensions mentioned previously, EE (9 items), D (5 items) and PA (8 items). The first version, aimed at health professionals, was called the MBI-Human Services Survey (MBI-HSS) and the second was aimed at education professionals, MBI-Educators' Survey (MBI-ES). The MBI-General Survey (MBI-GS), a reduced version of 16 items to measure Burnout in any profession (Maslach, Jackson, and Leiter, 1996), was published at the same time.

The MBI is one of the instruments most widely used to evaluate the syndrome (Worley, Vassar, Wheeler, and Barnes, 2008) and it has been adap-

ted to different languages (Kokkinos, 2006), including Spanish (Seisdedos, 1997). The generalized use of the questionnaire confirms, should that be necessary, that its validity and reliability are adequate. There have been numerous studies evaluating the MBI construct; in this line of research, the meta-analytic study undertaken by Worley *et al.* (2008) is of great interest.

Since the publication of the questionnaire, a level of alpha values has been found that oscillates between .81 and .92 (.89 in the original validation) in the *Emotional exhaustion* dimension; the internal consistency level in *Personal accomplishment* oscillates between .50 and .86, and in *Depersonalisation* between .57 and .82, the values proposed initially by the authors of the MBI in these two dimensions being .74 and .77 (Aluja, Blanch, and García, 2005; Gil-Monte and Peiró, 1999; Kantas and Vassilaki, 1997; Kim and Ji, 2009; Maslach and Jackson, 1981; Richardsen and Martinussen, 2005). Some empirical studies report on the reliability values of the test in the sample used; however, generally this information is omitted when applying the instrument to a specific sample (Vacha-Haase, 1998). The researcher assumes the reliability obtained in previous applications, usually that presented in the manual (reliability induction), wrongly believing that the parameter is constant and does not depend on the characteristics of the sample to which it is being applied (Crocker and Algina, 1986; Thompson and Vacha-Haase, 2000; Vacha-Haase, Kogan, and Thompson, 2000). This process of reliability induction would be adequate assuming that the target sample of the study was confirmed as similar in composition and variability to that of the reference sample; but this check is not usually carried out (Vacha-Haase *et al.* 2000). As a consequence, both the statistical power and the estimates of effect-sizes obtained can be wrongly interpreted (Wilkinson and APA Task Force on Statistical Inference, 1999).

Consequently, the existence of different versions and adaptations to different languages of the MBI, together with its broad use in diverse settings and populations, as well as the variability in the reliability values found, make it necessary to verify whether the reliability of the scores can be generalized across research studies. To achieve this task, a generalization study (RG) (Vacha-Haase, 1998), a very useful meta-analytic technique for the revision, integration and analysis of research results (Botella and Gambara, 2006; Montero and León, 2007; Sánchez-Meca and Botella, 2010), was carried out. More specifically, the goals of the present research were: a) to check whether reliability estimates can be generalized in the studies where

MBI has been applied and, if variability in the estimates of this parameter is found, b) to examine the variables that explain its heterogeneity, in each of the dimensions.

Method

Search of studies and selection criteria

The search of empirical studies was carried out in the following databases: Scopus, PsycARTICLES, Proquest, CSA, PsycINFO, Dialnet and Psicodoc, complemented by the search engine Google Academic. The words used were "Maslach Burnout Inventory" and "MBI", combined with the terms "reliability", "accuracy", "psychometric properties", "meta-analysis" and "review". The search was carried out during March and April of 2009, without any time restriction. To access "grey literature", the following bases were used: Web of Science, TESEO, System for Information on Grey Literature in Europe and National Technical Information Service.

The criteria for inclusion were: a) they must be empirical studies with MBI and b) they must report on the alpha coefficient of the sample. The search produced 84 studies, of which 24 in which MBI-GS was applied, 6 in which only a part of the test was applied and 9 in which reliability was not reported on were eliminated. In the end, 45 studies were used, the unit of analysis being the sample of participants and not the number of investigations (51 samples and 25,337 participants).

Codification of the variables

The moderating variables included to examine their influence in the reliability estimate of the MBI were:

1) Age: mean and standard deviation of age in the sample.
2) Sex: percentage of men in each sample.
3) Type of population: 1, health-related workers; 2, teachers; 3, other workers; 4, workers of various occupations.
4) Versions: 1, MBI-HSS; 2, MBI-ES.
5) Size: sample size.
6) Type of test: 1, original; 2, adaptation.
7) Language of the test: 1, English; 2, Spanish; 3, others.
8) Average scores: average of the scores obtained in the three dimensions.

9) SD of the scores: standard deviation of the scores obtained in each dimension.
10) Sampling: 1, probabilistic; 2; non-probabilistic.
11) Purpose of the study: 1, psychometric; 2, substantive characteristic; 3, both purposes.
12) Type of publication: 1, article in a journal with impact; 2, article in a journal without impact; 3, article in a journal not indexed in the databases; 4, minutes of a conference; 5, Internet document that includes none of the previous characteristics.
13) Countries where the study was undertaken: 1, North America; 2, Europe; 3, others.

The differentiation between journals with and without impact (12) was made in order to check whether differences exist according to the quality of the journal at the time when the alpha coefficients obtained in the work sample are incorporated, since it is to be expected that journals with impact have stricter selection criteria for articles.

The dependent variable appears in the codification manual below the α tag (value of Cronbach's alpha coefficient for each of the dimensions: EE, PA and D). The table with the data from the RG study is shown in the Appendix.

To evaluate the degree of reliability of the codification process, help from independent codifiers was sought. The degree of agreement between judges, evaluated as a percentage, was very high (94%).

Statistical analysis

The alpha coefficients were transformed into *T* scores to normalise the reliability estimates (Hakstian and Whalen, 1976) in such a way that, $T_i = (1-r_i)^{1/3}$ being the transformed coefficient and r_i the alpha coefficient. Each reliability estimate transformed by the inverse of its sample variance was adjusted to reflect its degree of precision (Sánchez-Meca and López-Pina, 2008).

The *Q* test (Hedges and Olkin, 1985) was used to examine whether the reliability coefficients were homogeneous, this being complemented by the I^2 index (Higgins and Thompson, 2002), following the recommendations of Huedo-Medina, Sánchez-Meca, Marín-Martínez, and Botella (2006).

After analysing the homogeneity of the coefficients and finding that the variances were heterogeneous, the average reliability was calculated, assu-

ming the model of random effects, since the estimation of between-studies variance was greater than zero (Sánchez-Meca, Marín-Martínez, and Huedo-Medina, 2006).

The effect of the moderating variables on the variability of the reliability estimates was evaluated by means of ANOVAs for categorical variables and regression models for continuous variables, assuming a model of mixed effects. Finally, a multiple regression model was used, with the aim of presenting an explanatory model that included the most relevant variables for the prediction of reliability estimates. The statistical package SPSS 15.0.1 and Excel 2007 were used to carry out the analysis. The recommendations put forward by Botella and Gambara (2002) were followed.

Results

Description of the characteristics of the studies

Of the 45 investigations included in the study, 60.9% were published between 2004 and 2009, 54.7% in journals with an impact index. The average age of participants was 38.32 (*SD* = 2.23) and 46.5% were in the teaching profession. The average percentage of males was 40.37% (*SD* = 22.29%). The version of the MBI most used was the MBI-HSS (55.1%). With regard to the type of test, 50% used the original, administered in English. A non-probabalistic sampling was utilised in 70.4% of the samples and in 55.2% the aim was to evaluate the psychometric properties of the MBI.

Estimate of average reliability

The average reliability obtained with the alpha coefficients without any weighting factor was, for EE (α= .87 and *SD* = 0.05), D (α= .70 and *SD* = 0.09) and PA (α= .76 and *SD* = 0.08). The minimum coefficients were: .66 (EE), .43 (D), and .49 (PA), and the maximum: .95 (EE), .83 (D) and .94 (PA) (see Figure 1 to examine the distribution of the alpha coefficients of each dimension).

Emotional exhaustion stem and leaf plot

Frequency	Stem and leaf
3.00	Extremes (=< .77)
1.00	7. 9
5.00	8. 12334
27.00	8. 555555667777778888888999999
14.00	9. 00000011222233
1.00	9. 5

Stem width: .10
Each leaf: 1 case(s)

Depersonalization stem and leaf plot

Frequency	Stem and leaf
1.00	Extremes (=< .43)
3.00	5. 033
3.00	5. 778
8.00	6. 02333444
7.00	6. 5566788
11.00	7. 01122222444
11.00	7. 56667899999
7.00	8. 0001233

Stem width: .10
Each leaf: 1 case(s)

Personal Accomplishment stem and leaf plot

Frequency	Stem and leaf
3.00	Extremes (=< .55)
2.00	6. 99
12.00	7. 001112223444
17.00	7. 56677777888999999
12.00	8. 000111112344
4.00	8. 5556
1.00	Extremes (>= .94)

Stem width: .10
Each leaf: 1 case(s)

FIGURE 1. Distribution of the alpha coefficient for each dimension.

The result of the Q test leads us to reject the hypothesis of homogeneity in the reliability estimates for the three dimensions, $Q_{EE(50)}$ = 740.99, p < .01; $Q_{D(50)}$ = 940.26, $p < .01$; and $Q_{PA(50}$ = 1,233.72, $p < .01$. The I^2 values indicate that in 93.25% in the case of EE, 94.68% in that of D and 95.95% in that of PA, the variability is due to the fact that the reliability estimates are heterogeneous.

Applying the model of random effects and weighting the coefficient transformed to *T* scores by the inverse of the sum of the sampling variance of the statistic and the inter-study variance (τ^2) (Sánchez-Meca and López-Pina, 2008), the average reliability found was EE (α= .88 and *SD* = 0.05), D (α= .71 and *SD* = 0.09) and PA (α= .78 and *SD* = 0.08). The confidence interval at 95% for the average reliability studied in the three dimensions was: EE (.87-.89), D (.68-.73), and PA (.75-.79).

Relation of moderating variables to reliability estimate

In the EE dimension it was found that the dispersion of scores was directly associated with the average reliability estimate, which explains the higher proportion of variance (R^2_{adj} = .45) and the model was found to be well specified, as indicated by the lack of statistical significance of the Q_E test; that is to say, a greater variability in the scores, a higher reliability estimate (the negative sign of the regression coefficient b = -0.031 results from the *T* transformation used, which inverts the order of the original coefficients and indicates the existence of a positive relation) (see Table 1). The second variable that explained the higher proportion of variance was the country where the study was undertaken (η^2 = .31), the studies from North America being those obtaining the highest reliability coefficient (α= .91) (see Table 2). Other variables that obtained significant differences with an explained variance of less than 20% were: (1) type of sample (η^2 = .19), with a higher estimate for the studies undertaken with probabilistic sampling (α= .90); (2) average age of participants (R^2_{adj} = .16), where the higher the average age of the sample, the higher the reliability estimate obtained; (3) language of the MBI (η^2 = .14), the studies where the MBI was administered in English being those that obtained a higher reliability coefficient (α= .90); (4) type of MBI (η^2 = .12), the studies that applied the original MBI demonstrating a greater reliability (α= .90). The variable version of MBI presented a low proportion of explained variance (η^2 = .06), the reliability estimate being

higher for the studies that used MBI-ES (α= .89) (see Tables 1 and 2). The other variables evaluated did not reach statistical significance.

TABLE 1. Simple regression models for the continuous variables in EE.

Moderator variable	*K*	*b*	Q_R	Q_E	R^2_{adj}
$\overline{X}$ Age	34	-0.005	11.01**	48.80*	.16
SD Scores	30	-0.031	26.85**	30.10	.45

Notes. b = Unstandardized regression coefficient; Q_R = Weighted regression sum of squares with 1 degree of freedom to assess the model fitting; Q_E = Weighted error sum of squares with k - 2 degrees of freedom to assess the model misspecification; *$p < .05$; **$p < .01$.

TABLE 2. ANOVAs for the categorical moderator variables in EE.

Moderator variable	*K*	$\overline{\alpha}$	*95% CI*	Q_w	I^2	η^2
Version MBI				Q_B = 4.98*		.06
HSS	28	.88	[.87, .88]	42.99*	37.2	
ES	22	.89	[.89, .89]	28.66	26.7	
Type MBI				Q_B = 9.51**		.12
Original	26	.90	[.90, .91]	48.84**	48.81	
Adaptation	24	.87	[.87, .87]	18.27	0	
Language MBI				Q_B = 10.74**		.14
English	26	.90	[.90, .91]	48.84**	48.8	
Spanish	12	.87	[.87, .88]	5.35	0	
Others	11	.87	[.87, .88]	11.69	14.4	
Sampling				Q_B = 14.26**		.19
Probabilistic	13	.90	[.90, .90]	11.37	0	
Non probabilistic	37	.87	[.87, .87]	50.99	29.4	
Countries				Q_B = 23.68**		.31
North America	19	.91	[.90, .91]	13.21	0	
Europe	18	.87	[.87, .88]	14.16	0	
Others	13	.86	[.85, .87]	25.56*	53.1	

Notes. $\overline{\alpha}$ = Weighted average reliability estimate in terms of alpha coefficient; Q_w = Within-category heterogeneity statistic with k – 1 degrees of freedom; Q_B = statistic for testing the influence of the moderator variables on the score reliability estimates; *$p < .05$; **$p < .01$.

In the D dimension (see Table 3), the moderating variable that explained the highest proportion of variance of the average reliability estimate was language of the MBI (η^2 = .28), this reaching a higher reliability coefficient in the studies that applied the MBI in English (α= .76). The second variable that explained the highest proportion of variance was type of MBI (η^2 = .23), the studies applying the original MBI obtaining a higher reliability coefficient (α= .76). The variable country where the study was undertaken also explained an average proportion of the reliability estimate (η^2 = .20). The studies carried out in North American countries showed a higher reliability

coefficient (α= .76). No significant differences were obtained in any of the other modulating variables examined.

TABLE 3. ANOVAs for the categorical moderator variables in D.

Moderator variable	*K*	$\overline{\alpha}$	*95% CI*	Q_w	I^2	η^2
Type MBI				Q_B = 10.29**		.23
Original	26	.76	[.75, .77]	10.71	0	
Adaptation	24	.64	[.63, .65]	23.11	0.48	
Language MBI				Q_B = 12.39**		.28
English	26	.76	[.75, .77]	10.71	0	
Spanish	12	.60	[.58, .61]	12.5	12	
Others	11	.69	[.67, .70]	8.55	0	
Countries				Q_B = 8.72*		.20
North America	19	.76	[.75, .77]	8.33	0	
Europe	18	.68	[.67, .69]	12.51	0	
Others	13	.63	[.61, .65]	14.55	17.5	

Notes. $\overline{\alpha}$ = Weighted average reliability estimate in terms of alpha coefficient; Q_w = Within-category heterogeneity statistic with k – 1 degrees of freedom; Q_B = statistic for testing the influence of the moderator variables on the score reliability estimates; *p < .05; **p < .01.

In the PA dimension of work, no significant differences were found in any of the moderating variables evaluated.

Explanatory model

The explanatory model proposed in the EE version includes the moderating variables: dispersion of scores, sample and country where the study was carried out. This model turned out to be very significant, $Q_{R(4)}$ = 30.41, p < .001, explaining 57% of the variance (R_{2adj} = .57) and being well specified, as indicated by the lack of statistical significance of the test Q_E, $Q_{E(47)}$ = 17.56.

In the D dimension, the variables making up the multiple regression model were language of the MBI and type of MBI. The distribution of scores was also included, following the recommendations of Rodríguez and Maeda (2006), although it did not reach statistical significance in the simple regression model. This model explains 35% of the variance (R_{2adj} = .35), reaching statistical significance, $Q_{R(4)}$ = 9.18, p < .05, and showing an adequate specification as indicated by the lack of statistical significance of the test Q_E, $Q_{E(47)}$ = 12.9.

Discussion

In the present research, the average reliability estimates in each dimension of the MBI (EE, D and PA) were calculated and the moderating variables

that could explain the variability found were studied. To do this, empirical research studies that applied the MBI and reported on the alpha reliability coefficient of the sample were reviewed. Of a total of 84 studies found, 51 samples obtained from 45 investigations were analysed.

The average alpha coefficient obtained across the 51 reliability estimations was .88 for the EE dimension, .71 for D and .78 for PA. The highest average reliability estimate was obtained in the EE dimension, followed by PA and finally in the D dimension. The average reliability values obtained are lower than those reported by Maslach and Jackson (1981) for EE (.89) and D (.77) and higher for PA (.74).

The reliability estimates presented a high level of heterogeneity in the three dimensions (the I_2 index was 93.7% for EE, 95.5% for D and 96.3% for PA). Consequently the reliability of the scores in each of the dimensions of the MBI cannot be generalized to the different populations and contexts represented in the meta-analysis. The practice of reliability induction when this questionnaire is applied is inadvisable since it could have an influence on interpretations of the statistical power of the hypothesis test and the estimation of effect size (Wilkinson and APA Task Force on Statistical Inference, 1999).

When variability was found in the reliability estimates in the three dimensions of the MBI, the moderating variables that could explain this heterogeneity were examined. In the EE dimension, the variables that influenced the reliability estimate were distribution of scores, country where the study was undertaken, type of sample, average age of participants, language, type of MBI and version of MBI. The reliability estimates were higher in the following situations: a) the greater the variability of scores; b) studies undertaken in North America; c) investigations conducted using a probabilistic sampling; d) the higher the average age of the sample; e) studies that administered the MBI in English; f) investigations that applied the original MBI and g) studies that used the MBI-ES. Thus, there are several moderating variables that explain, in part, the heterogeneity found in the reliability estimates, leading us to suggest an explanatory model of this variability in which the most relevant moderating variables were dispersion of scores, sample and country where the study was undertaken.

In the D dimension, the moderating variables that explained the variability in reliability estimation were language and type of MBI, as well as the country where the study was undertaken. The reliability estimates were

higher when the studies applied the MBI in English, in the original version and when it was carried out in North America. These variables, with the exception of country where the study was undertaken, together with the distribution of scores, were examined jointly in a possible explanatory model, which concluded that all these influenced alpha reliability estimates.

With regard to the PA dimension, the moderating variables evaluated did not explain the heterogeneity of the reliability estimates. Without a doubt, the variable dispersion of scores also explained an acceptable proportion of variance, although no significant differences were obtained.

The results are congruent with the fact that reliability is a property of the scores obtained when a measuring instrument is applied in a specific sample, and is not an intrinsic characteristic of it (Crocker and Algina, 1986; Thompson and Vacha-Haase, 2000). It would be inadvisable to continue with the practice of reliability induction when the MBI is administered: the reliability obtained in each application should be included in the empirical studies. Since the reliability estimates in the three dimensions of the MBI depend on the characteristics of the sample, wrong decisions could be made, for example, in the D dimension, the precision and interpretation of the results could vary according to whether an adaptation (α= .64) was applied or the original version of the instrument (α= .76).

In this study of RG, Cronbach's alpha, transformed to *T* scores in order to achieve a better approximation to normal distribution and to stabilise the variance, was used as reliability estimator (Sánchez-Meca and López-Pina, 2008). Future investigations could examine whether the results found in this study are corroborated when using other reliability estimators, although this would be very difficult, since reliability indicators other than the alpha coefficient are hardly ever used for a questionnaire.

To conclude, in the case of MBI, information as to the specific reliability of each dimension, according to the sample used, should always be included. If not, errors could be made when estimating the precision with which the measurement of Burnout is carried out.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Aluja, A., Blanch, A., and García, L.F. (2005). Dimensionality of the Maslach Burnout Inventory in school teachers: A study of several proposals. *European Journal of Psychological Assessment, 21*, 67-76.

*Bahner, A.D. and Berkel, L.A. (2007). Exploring burnout in batterer intervention programs. *Journal of Interpersonal Violence, 22*, 994-1008.

*Beckstead, J.W. (2002). Confirmatory factor analysis of the Maslach Burnout Inventory among Florida nurses. *International Journal of Nursing Studies, 39*, 785-792.

Bernaldo de Quirós-Aragón, M. and Labrador-Encinas, F. (2007). Evaluación del estrés laboral y burnout en los servicios de urgencia extrahospitalaria. *International Journal of Clinical and Health Psychology, 7*, 323-335.

*Boles, J.S., Dean, D.H., Ricks, J.M., Short, J.C., and Wang, G. (2000). The dimensionality of the Maslach Burnout Inventory across small business owners and educators. *Journal of Vocational Behaviour, 56*, 12-34.

Botella, J. and Gambara, H. (2002). *Qué es el meta-análisis*. Madrid: Biblioteca Nueva.

Botella, J. and Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical and Health Psychology, 6*, 425-440.

*Buzzetti, M.A. (2005). *Validación del Maslach Burnout Inventory (MBI) en dirigentes del colegio de profesores A. G. de Chile* (unpublished master´s thesis). Facultad de Ciencias Sociales, Universidad de Chile.

*Byrne, B.M. (1991). Burnout: Investigating the impact of background, variables for elementary, secondary, and university educators. *Teaching and Teacher Education, 7*, 197-209.

*Byrne, B.M. (1993). The Maslach Burnout Inventory: Testing for factorial validity and invariance across elementary, intermediate and secondary teachers. *Journal of Occupational and Organizational Psychology, 66*, 197-212.

*Carlotto, M.S. and Câmara, S.G. (2007). Propriedades psicométricas do Maslach Burnout Inventory em uma amostra multifuncional. *Estudos de Psicologia (Campinas), 24*, 325-332.

Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Holt, Rinehart, and Winston.

Freudenberger, H. (1974). Staff Burn-Out. *The Journal of Social Issues, 30*, 159-165.

*García, J.M., Herrero, S., and León, J.L. (2007). Validez factorial del Maslach Burnout Inventory (MBI) en una muestra de trabajadores del Hospital Psiquiátrico Penitenciario de Sevilla. *Apuntes de Psicología, 25*, 157-174.

*Garman, A.N., Corrigan, P.W., and Morris, S. (2002). Staff burnout and patient satisfaction evidence of relationships at the care unit level. *Journal of Occupational Health Psychology, 7*, 235-241.

*Gavrilovici, O. (2008). *Romanian teachers´ burnout and psychological and professional difficulties*. Retrieved April 22, 2009, from http://holon.ladipu.com/resources/15/Romanian%20teachers%27%20burnout%20and%20.pdf

*Genoud, P.A., Brodard, F., and Reicherts, M. (2009). Facteurs de stress et burnout chez les enseignants de l´ècole primaire. *Revue Europèenne de Psychologie Appliquée, 59*, 37-45.

*Gil-Monte, P.R. (2005). Factorial validity of the Maslach Burnout Inventory (MBI-HSS) among Spanish professionals. *Revista de Saúde Pública, 39*, 1-8.

*Gil-Monte, P.R., García-Juesas, J.A., and Caro, M. (2008). Influencia de la sobrecarga laboral y la autoeficacia sobre el síndrome de quemarse por el trabajo (Burnout) en profesionales de enfermería. *Revista Interamericana de Psicología, 42*, 113-118.

*Gil-Monte, P.R. and Peiró, J.M. (1999). Validez factorial del Maslach Burnout Inventory en una muestra multiocupacional. *Psicothema, 11*, 679-689.

*Gold, I. (1984). The factorial validity of the Maslach Burnout Inventory in a sample of California elementary and junior high school classroom teachers. *Educational and Psychological Measurement, 44*, 1009-1016.

*Grau, A., Flichtentrei, D., Suñer, R., Font-Mayolas, S., Prats, M., and Braga, F. (2008). El burnout percibido o sensación de estar quemado en profesionales sanitarios: prevalencia y factores asociados. *Información Psicológica, 91*, 64-79.

*Grayson, J.L. and Alvarez, H.K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education, 24*, 1349-1363.

Hakstian, A.R. and Whalen, T.E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika, 41*, 219-231.

*Hastings, R.P., Horne, S., and Mitchell, G. (2004). Burnout in direct care staff in intellectual disability services: A factor analytic study of the Maslach Burnout Inventory. *Journal of Intellectual Disability Research, 48*, 268-273.

Hedges, L.V. and Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

*Hernández-Martín, L., Fernández-Calvo, B., Ramos, F., and Contador, I. (2006). El síndrome de burnout en funcionarios de vigilancia de un centro penitenciario. *International Journal of Clinical and Health Psychology, 6*, 599-611.

Higgins, J.P.T. and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.

*Hogan, R.L. and McKnight, M.A. (2007). Exploring burnout among university online instructors: An initial investigation. *The Internet and Higher Education, 10*, 117-124.

Huedo-Medina, T.B., Sánchez-Meca, J., Marín-Martínez, F., and Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*, 193-206.

*Ioannou, I. and Kyriakides, L. (2006, October). *Structuring a model for the determinants of vocacional teacher burnout*. Paper presented at the Commonwealth Council for Educational Administration and Management, Nicosia.

Jenaro-Río, C., Flores-Robaina, N., and González-Gil, F. (2007). Síndrome de burnout y afrontamiento en trabajadores de acogimiento residencial de menores. *International Journal of Clinical and Health Psychology, 7*, 107-121.

*Kanste, O., Miettunen, J., and Kyngäs, H. (2006). Factor structure of the Maslach Burnout Inventory among Finnish nursing staff. *Nursing and Health Sciences, 8*, 201-207.

*Kantas, A. and Vassilaki, E. (1997). Burnout in Greek teachers: Main findings and validity of the Maslach Burnout Inventory. *Work and Stress, 11*, 94-100.

*Khani, A., Jaafarpour, M., and Jamshidbeigi, Y. (2008). The relationship between clinical supervision and burnout in the nurse´s job an Iranian study. *Journal of Clinical and Diagnostic Research, 2*, 913-918.

*Kim, H. and Ji, J. (2009). Factor structure and longitudinal invariance of the Maslach Burnout Inventory. *Research on Social Work Practice, 19*, 325-339.

*Kokkinos, C.M. (2006). Factor structure and psychometric properties of the Maslach Burnout Inventory- Educators Survey among elementary and secondary school teachers in Cyprus. *Stress and Health, 22*, 25-33.

*Koustelios, A. and Tsigilis, N. (2005). The relationships between burnout and job satisfaction among physical education teachers: A multivariate approach. *European Physical Education Review, 11*, 189-203.

*Lee, H., Song, R., Cho, Y.S., Lee, G.Z., and Daly, B. (2003). A comprehensive model for predicting burnout in Korean nurses. *Journal of Advanced Nursing, 44*, 534-545.

*Manso-Pinto, J.F. (2006). Estructura factorial del Maslach Burnout Inventory- version Human Services Survey- en Chile. *Revista Interamericana de Psicología, 40*, 115-118.

*Martínez-López, C. and López-Solache, G. (2005). Características del síndrome de burnout en un grupo de enfermeras mexicanas. *Archivos en Medicina Familiar, 7*, 6-9.

Maslach, C. and Jackson, S.E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior, 2*, 99-113.

Maslach, C., Jackson, S.E., and Leiter, M.P. (1996). *The Maslach Burnout Inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

*Meda, R.M., Moreno, B., Rodríguez, A., Morante, E., and Ortiz, G.R. (2008). Análisis factorial confirmatorio del MBI-HSS en una muestra de psicólogos mexicanos. *Psicología y Salud, 18*, 107-116.

Montero, I. and León, O.G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology, 7*, 847-862.

*Moreno, B., Meda, R.M., Rodríguez, A., Palomera, A., and Morales, M.M. (2006). El síndrome de burnout en una muestra de psicólogos mexicanos: prevalencia y factores sociodemográficos asociados. *Psicología y Salud, 16*, 5-13

*Naudé, J.L.P. and Rothmann, S. (2004). The validation of the Maslach Burnout Inventory- Human Services Survey for emergency medical technicians in Gauteng. *SA Journal of Industrial Psychology, 30*, 21-28.

*O´Brien, P., O´Keeffe, M., and Goddard, R. (2007, November). *Burnout confirmed as a viable explanation for beginning teacher attrition*. Paper session presented at the meeting of Australian Association for Research in Education Annual Conference, Fremantle.

*Oh, S-.H. and Lee, M. (2009). Examining the psychometric properties of the Maslach Burnout Inventory with a sample of child protective service workers in Korea. *Children and Youth Services Review, 31*, 206-210.

*Önder, Ç. and Basim, N. (2008). Examination of developmental models of occupational burnout using burnout profiles of nurses. *Journal of Advanced Nursing, 64*, 514-523.

*Powers, S. and Gose, K.F. (1986). Reliability and construct validity of the Maslach Burnout Inventory in a sample of university students. *Educational and Psychological Measurement, 46*, 251-255.

*Richardsen, A.M. and Martinussen, M. (2004). The Maslach Burnout Inventory: Factorial validity and consistency across occupational groups in Norway. *Journal of Occupational and Organizational Psychology, 77*, 377-384.

Richardsen, A.M. and Martinussen, M. (2005). Factorial validity and consistency of the MBI-GS across occupational groups in Norway. *International Journal of Stress Management, 12*, 289-297.

Rodríguez, M.C. and Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 3*, 306-322.

*Rohland, B.M., Kruse, G.R., and Rohrer, J.E. (2004). Validation of a single-item measure of burnout against the Maslach Burnout Inventory among physicians. *Stress and Health, 20*, 75-79.

Sánchez-Meca, J. and Botella, J. (2010). Revisiones sistemáticas y meta-análisis: herramientas para la práctica profesional. *Papeles del Psicólogo, 31*, 7-17.

Sánchez-Meca, J. and López-Pina, J.A. (2008). El enfoque meta-analítico de generalización de la fiabilidad. *Acción Psicológica, 5*, 37-64.

Sánchez-Meca, J., Marín-Martínez, F., and Huedo-Medina, T.B. (2006). Modelo de efectos fijos y modelo de efectos aleatorios. In J.L.R. Martín, A. Tobías, and T. Seoane (Eds.), *Revisiones sistemáticas en las ciencias de la vida* (pp. 189-206). Toledo: FISCAM.

*Schaufeli, W.B., Bakker, A.B., Hoogduin, K., Schaap, C., and Kladler, A. (2001). On the clinical validity of the Maslach Burnout Inventory and the Burnout Measure. *Psychology and Health, 16*, 565-582.

*Schaufeli, W.B., Daamen, J., and Van Mierlo, H. (1994). Burnout among Dutch teachers: An MBI-validity study. *Educational and Psychological Measurement, 54*, 803-812.

*Schaufeli, W.B. and Van Dierendonck, D. (1993). The construct validity of two burnout measures. *Journal of Organizational Behavior, 14*, 631-647.

Seisdedos, N. (1997). *MBI. Inventario "Burnout" de Maslach: manual*. Madrid: TEA.

Topa-Cantisano, G. and Morales-Dominguez, J.F. (2007). Burnout e identificación con el grupo: el papel del apoyo social en un modelo de ecuaciones estructurales. *International Journal of Clinical and Health Psychology, 7*, 337-348.

Thompson, B. and Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174–195.

Vacha-Haase, T. (1998). Reliability generalization: exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20.

Vacha-Haase, T., Kogan, L.R., and Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*, 509-522.

*Viloria, H. and Paredes, M. (2002). Estudio del síndrome de burnout o desgaste profesional en los profesores de la Universidad de Los Andes. *Educere, 17*, 29-36.

*Walter, J.M., Van Lunen, B.L., Walker, S.E., Ismaeli, Z.C., and Oñate, J.A. (2009). An assessment of burnout in undergraduate athletic training education program directors. *Journal of Athletic Training, 44*, 190-196.

Wilkinson, L. and APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journal: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Worley, J.A., Vassar, M., Wheeler, D.L., and Barnes, L.L.B. (2008). Factor structure of scores from the Maslach Burnout Inventory: A review and meta-analysis of 45 exploratory and confirmatory factor-analytic studies. *Educational and Psychological Measurement, 68,* 797-823.

Received April 20, 2010
Accepted December 12, 2010

APPENDIX. Data for each sample included in the RG study.

Study	*Age*	*σAge*	*Sex*	*Population*	*Version*	*Size*	*α AE*	*α D*	*α RP*	*Test*	*Language*	$\bar{X}_{AE}$	$\bar{X}_{D}$	$\bar{X}_{RP}$	*σAE*	*σD*	*σRP*	*Sampling*	*Purpose*	*Publication*	*Countries*
Aluja *et al.* (2005)	38.34	8.19	28.68	2	2	631	.83	.50	.79	2	3	17.90	4.23	36.54	8.02	4.31	7.29	2	3	1	2
Bahner *et al.* (2007)	43.00	11.77	44.35	3	1	115	.88	.64	.49	1	1	15.86	6.64	34.27	8.80	4.69	5.93	2	2	1	1
Beckstead (2002)	36.50	-	85.43	1	1	151	.88	.80	.75	1	1	22.81	6.97	38.81	11.08	5.77	9.01	2	1	1	1
Boles *et al.*(2000a)	43.00	-	30.05	2	2	183	.89	.80	.76	1	1	-	-	-	-	-	-	2	1	1	1
Boles *et al.* (2000b)	45.00	-	59.26	3	1	162	.90	.70	.78	1	1	-	-	-	-	-	-	2	1	1	1
Buzzetti (2005)	-	-	-	2	2	98	.92	.76	.55	2	2	-	-	-	-	-	-	2	1	5	3
Byrne (1991a)	-	-	83.67	2	2	98	.90	.64	.69	1	1	21.59	5.24	40.28	11.71	3.63	6.00	1	2	1	1
Byrne (1991b)	-	-	46.01	2	2	163	.92	.83	.77	1	1	23.24	7.41	39.19	11.78	6.41	6.29	1	2	1	1
Byrne (1991c)	-	-	62.35	2	2	162	.89	.71	.84	1	1	21.27	6.64	35.88	10.71	5.15	8.19	1	2	1	1
Byrne (1991d)	-	-	47.49	2	2	219	.89	.74	.83	1	1	17.80	7.01	36.90	10.46	6.51	7.64	1	2	1	1
Byrne (1993a)	-	-	42.02	2	2	1159	.91	.74	.79	1	1	22.54	5.37	40.48	11.26	5.10	5.73	1	1	1	1
Byrne (1993b)	-	-	67.53	2	2	388	.91	.72	.80	1	1	23.71	7.07	39.21	11.15	5.45	6.29	1	1	1	1
Byrne (1993c)	-	-	47.4	2	2	1384	.90	.74	.80	1	1	22.56	6.15	38.43	10.76	5.16	6.58	1	1	1	1
Carlotto *et al.* (2007)	-	-	44.06	4	1	665	.85	.65	.94	2	3	-	-	-	-	-	-	2	1	2	3
García *et al.* (2007)	41.61	-	80.15	4	1	136	.88	.79	.81	2	2	22.68	9.58	28.76	11.71	7.51	9.12	2	1	2	2
Garman *et al.* (2002)	44.00	10.00	24.92	1	1	333	.95	.79	.81	1	1	17.20	5.40	37.40	10.90	5.20	7.90	-	2	2	1
Gavrilovici (2008)	40.60	11.40	23.6	2	1	178	.79	.62	.78	2	3	28.34	9.30	22.45	9.33	4.34	7.71	1	3	5	2
Genoud *et*	-	-	24.02	2	-	787	.88	.65	.79	2	3	-	-	-	-	-	-	2	2	2	2

al. (2009)																									
Gil-Monte *et al.* (1999)	33.90	8.12	35.6	4	1	559	.87	.57	.72	2	2	20.34	6.04	36.39	11.03	5.18	6.88	2	1	1	2				
Gil-Monte (2005)	32.50	-	40.85	4	1	705	.85	.58	.71	2	2	20.68	6.46	35.77	10.81	5.35	7.39	2	1	2	2				
Gil-Monte *et al.* (2008)	38.60	8.71	16.67	1	1	714	.89	.64	.79	2	2	17.87	5.47	36.01	10.78	4.81	7.33	2	2	2	2				
Gold (1984)	-	-	19.05	2	2	462	.88	.72	.74	1	1	-	-	-	-	-	-	2	1	1	1				
Grau *et al.* (2008)	42.41	10.90	52.68	1	1	6150	.87	.57	.72	2	2	26.07	8.02	37.17	-	-	-	2	3	2	3				
Grayson *et al.* (2008)	42.34	11.59	26.25	2	2	320	.90	.79	.71	1	1	22.94	5.71	37.84	10.54	5.00	7.03	2	2	1	1				
Hastings *et al.* (2004)	33.09	9.82	37.5	3	1	184	.87	.68	.76	1	1	-	-	-	-	-	-	2	1	1	2				
Hernández-Martín *et al.* (2006)	40.68	5.18	89.47	3	1	133	.85	.66	.70	2	2	24.67	12.60	21.24	11.81	6.86	7.95	2	3	1	2				
Hogan *et al.* (2007)	-	-	60.53	2	2	76	.93	.72	.73	1	1	22.08	9.81	46.91	11.59	4.68	6.62	2	1	2	1				
Ioannou *et al.* (2006)	-	-	37.93	2	2	319	.84	.75	.77	2	-	-	-	-	-	-	-	2	2	4	2				
Kanste *et al.* (2006)	43.70	9.00	5.58	1	1	627	.87	.83	.81	2	3	-	-	-	-	-	-	1	1	2	2				
Kantas *et al.* (1997)	-	-	53.64	2	2	220	.81	.60	.86	2	3	18.80	3.90	38.60	10.00	4.00	8.00	2	1	1	2				
Khani *et al.* (2008)	32.00	6.00	29.17	1	1	120	.87	.72	.70	-	-	22.64	7.60	33.24	11.54	4.35	5.30	1	2	2	3				
Kim *et al.* (2009)	45.20	12.30	19.16	3	1	475	.92	.77	.80	1	1	-	-	-	-	-	-	1	1	1	1				
Kokkinos (2006)	37.51	10.60	28.15	2	2	771	.85	.63	.79	2	3	27.37	4.42	38.13	10.91	4.23	6.94	2	1	2	2				
Koustelios *et al.* (2005)	37.64	3.68	62.35	2	2	170	.76	.76	.82	2	3	15.24	3.52	37.84	8.35	3.36	7.66	2	2	1	2				
Lee *et al.* (2003)	30.00	-	-	1	1	178	.66	.72	.85	1	1	26.37	9.89	27.32	6.61	4.13	6.67	2	2	1	3				

Manso-Pinto (2006)	36.50	7.20	15.48	3	1	155	.82	.80	.85	2	2	-	-	-	-	-	-	1	2	3					
Martinez-Lopez *et al.* (2005)	-	-	11.41	1	1	622	.88	.43	.77	2	2	-	-	-	-	-	-	2	2	2	3				
Meda *et al.* (2008)	37.97	10.30	33.92	1	1	454	.86	.53	.81	2	2	13.46	-	-	9.39	4.20	6.67	2	1	2	3				
Moreno *et al.* (2006)	37.97	10.30	33.92	1	1	454	.87	.53	.81	2	2	-	-	-	-	-	-	2	2	3	3				
Naudé *et al.* (2004)	33.13	-	77.04	1	1	318	.77	.68	.78	1	1	15.84	11.03	20.69	6.72	5.55	5.94	2	1	2	3				
O'Brien *et al.* (2007)	27.46	7.32	20.41	2	2	98	.92	.79	.79	1	1	-	-	-	-	-	-	2	2	4	3				
Oh *et al.* (2009)	30.04	5.98	39.36	3	1	249	.83	.78	.84	1	1	-	-	-	-	-	-	2	1	1	3				
Onder *et al.* (2008)	31.18	-	-	1	2	248	.89	.76	.74	2	3	-	-	-	-	-	-	2	2	1	2				
Powers *et al.* (1986)	-	-	34.72	2	1	72	.86	.63	.72	1	1	21.74	12.39	39.71	11.54	7.12	7.52	2	1	1	1				
Richardsen *et al.* (2004)	40.80	10.70	19.56	4	1	1590	.90	.66	.77	2	3	-	-	-	-	-	-	1	1	1	2				
Rohland *et al.* (2004)	44.00	-	74.59	1	1	307	.85	.82	.52	1	1	22.00	8.20	40.10	11.00	6.10	5.80	2	1	2	1				
Schaufeli *et al.* (1993)	33.10	7.70	41.08	1	1	667	.88	.71	.77	2	3	16.68	5.94	32.41	8.45	3.89	4.91	2	1	1	2				
Schaufeli *et al.* (1994)	45.90	7.20	52.45	2	2	326	.93	.81	.85	1	1	-	-	-	-	-	-	2	1	1	1				
Schaufeli *et al.* (2001)	42.00	9.00	59.71	4	1	139	.89	.67	.74	1	1	28.53	9.24	27.04	10.07	4.89	5.68	2	1	1	2				
Viloria *et al.* (2002)	-	-	65.46	2	2	194	.85	.63	.69	2	2	-	-	-	-	-	-	1	2	2	3				
Walter *et al.* (2009)	41.19	7.99	41.77	2	2	249	.90	.79	.71	1	1	-	-	-	-	-	-	2	2	2	-				

Note. a, b, c, and d are different samples in the same study.