BMJ Open

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031859
Article Type:	Research
Date Submitted by the Author:	22-May-2019
Complete List of Authors:	Rodrigo, Maria F.; University of Valencia Molina, J. Gabriel; University of Valencia Losilla, Josep-Maria; Autonomous University of Barcelona Vives, Jaume; Autonomous University of Barcelona
Keywords:	psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

Title

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12)

2. Authors and affiliation

Rodrigo, Maria F.
Department of Research Methods in Psychology
University of Valencia
Valencia (Spain)

Molina, J. Gabriel †
Department of Research Methods in Psychology
University of Valencia
Valencia (Spain)

Losilla, Josep-Maria
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Barcelona (Spain)

Vives, Jaume
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Barcelona (Spain)

†Deceased 26 September 2014

3. Word count, excluding title page, abstract, references, figures and table: 3086

4. Authors' statements

All authors have agreed to authorship in the indicated order.
All authors declare that this paper is an original unpublished work and it is not being submitted elsewhere.
All authors do not have any financial interests that might be interpreted as influencing the research, and APA ethical standard were followed in the conduct of the study.
The research was not submitted to approval by an institutional review board since this is not a requirement at our universities for this type of study.

5. Contact information for the corresponding author:

Jaume Vives
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Edifici B - Despatx B5b/081
Carrer de la Fortuna / Carrer de Ca n'Altayó. Campus de la UAB
08193 Bellaterra (Cerdanyola del Vallès). Barcelona (Spain)
E-mail: jaume.vives@uab.cat
Phone: +34 93 581 23 31

Abstract

Objective. Recent studies into the factorial structure of the 12–item version of the General Health Questionnaire (GHQ–12) have shown that it was best represented by a single substantive factor when method effects associated with negatively worded (NW) items are considered. The purpose of the present study was to examine the presence of method effects, and their relationships with demographic covariates, associated with positively worded (PW) and/or NW items.

Method. The current work compared a comprehensive set of confirmatory factor models, including method effects associated with PW and/or NW items with QHQ-12 responses using a random sample of 3050 workers.

Results. A confirmatory factor analysis showed that the best–fitting model was a unidimensional model with two additional method factors associated with PW and NW items. Furthermore, structural equation modeling revealed that method effects were differentially related to both the sex and educational level of the respondents.

Conclusion. Individual differences related to sex and educational level can help to identify respondents who are prone to answering PW and NW items differently. Consequently, is desirable that both the constructs of interest as well as the effects of method factors are considered in SEM models as a means of avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

Keywords

psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

Strengths and limitations of this study

Strengths

- Sampling quality: A random and large representative sample of workers and face-to-face administration by professional interviewers.
- To compare confirmatory models for positively and/or negatively worded items and using two different parameterizations.
- To study demographic correlates of wording effects.

Limitations

- The different response scale used for the NW items and the PW items in the questionnaire could be a confusion variable.
- The results might not be generalized to other specific populations as, for example, adolescents and elderly retired people.

Introduction

Originally developed by Goldberg[1], the General Health Questionnaire (GHQ) has been widely used as a screening instrument for measuring General Psychological Health (GPH) in both community and non–psychiatric clinical settings[2]. The shortest 12–item version (GHQ–12) is the most popular and has been employed on different settings and in several countries, as well as part of multiple major national health, social wellbeing and occupational surveys, achieving results which underline the fact that it is highly reliable and valid[3–11].

Despite its broad application, the factor structure underlying the responses to the GHQ–12 remains a controversial issue. In this sense, although the GHQ–12 was originally developed as a unidimensional scale, this one–factor latent structure has found little empirical support and some alternative multidimensional models, have been proposed as more appropriate. Thus, the one with the most empirical support is the three–factor model proposed by Graetz[12][5,13,22,14–21]. It is important to note that the 6 positively worded (PW) items make up the first factor, whereas the other two factors are made up of the 6 negatively worded (NW) items (see Figure 1, Model 3). On the other hand, the bidimensional model, where the 6 NW and the 6 PW items in the GHQ–12 are grouped into two factors (see Figure 1, Model 2), has also obtained wide support, especially in studies based on exploratory factor analysis[5,10,23–28]. The arguments against these models and in favor of the unidimensional solution are the high correlations between the factors[13] and the low discriminant validity of the factor scores derived from these models[16,29,30].

As Hankins[31] points out, multifactor models may just be the resulting artifact of the inclusion of PW and NW items in the questionnaire and so, the controversy about the factorial structure of the GHQ–12 might relate to the effect of item wording on subjects' response patterns as part of a more general category called 'method[32,33]. Hankins[31] found that, after modeling the wording effects for the NW items, the unidimensional model fitted better than both the two–factor model (NW vs. PW items) and Graetz's three–factor model. Other studies have called into question the substantive meaning of the GHQ–12 multifactor solutions, suggesting that they might just be an

artifact due to the wording effects associated with NW items[29,30,34–40]. See Molina et al.[36] for a deeper review about the dimensionality of GHQ-12.

Some studies about other instruments, however, suggested not only considering the wording effects for the NW items but also for the PW items[41,42]. Regarding GHQ-12, only a recent analytical factor meta-analysis modelled the presence of method effects for negatively and positively worded items concluding that positively keyed items explained incremental variance beyond a general mental health factor[43].

Another source of variability in the results about the factor structure of the GHQ-12 could come from the statistical control of method biases, that has been mainly achieved through the correlated traits–correlated methods (CTCM) and the correlated traits–correlated uniquenesses (CTCU) confirmatory factor analysis models. Both procedures have been used in GHQ–12, to deal with method effects applying the CTCM model[i.e., 30,44] the CTCU model[i.e., 29,31,39,40], or both CTCM and CTCU[i.e., 34–37].

To date, we have not found any studies about GHQ-12 that analyze the wording effects associated with either PW items alone, nor with NW and PW items simultaneously, comparing both CTCU and CTCM models. So, this work extends the previous work by Molina et al.[36], which compares the fit of the unidimensional model, the multifactor models and the CTCM and CTCU unidimensional models with method effects for only the NW items. To clarify this work, Figure 1 (Model 1 to Model 6) shows the 6 CFA models that we consider here in order to test the potential method effects associated with either the PW items (Models 3 and 4) or both the NW and PW items (Models 5 and 6). Models 1 and 2 were the best fitted models in Molina et al.[36] and are the base-models for this study. Three of these models are CTCU models (Models 1, 3 and 5), whereas the other three are CTCM models (Models 2, 4 and 6). As stressed by Marsh et al.[45], it becomes necessary to consider this comprehensive set of competing models to determine the relative importance and substantive nature of the method effects.

Figure 1

Finally, there has been some research carried out into the demographic correlates of method effects, such as sex[46–50], age[48,51] or educational level[41,52]. With respect to the GHQ–12, to date, we have not found any studies that analyze demographic correlates of method effects.

Building on the previous studies, the first aim of this study was to overcome the limitation pointed out in Molina et al.[36] and examine method effects associated with both positive and negative wording. The second aim was to further understand the meaning of the method factors; therefore, we evaluated the relationships between the method factors and three covariates (i.e., the sex, age, and educational level) in the framework of a structural equation model (SEM).

Method

Participants

The data used in this study came from the Second Catalonian Survey of Working Conditions[53] and were based on a representative random sample of all employees living in Catalonia (Spain). Data were collected by professional interviewers in private households. The sample comprised a total of 3,050 participants who responded to the GHQ–12 included in the survey (55.4% men and 44.6% women) with a mean age of 40.46 years (*SD* = 11.19; range from 17 to 82).

Measures

The GHQ–12 is a self–report scale that contains 6 PW items (e.g. “Have you been able to face up to problems?”) and 6 NW items (e.g. “Have you been losing confidence in yourself?”). The GHQ–12 was validated in Spain by Lobo and Muñoz[54]. Table 1 shows the statements of these items in the same order as they were presented in the survey. It must be noted that the GHQ–12 has a different response scale for the PW items (i.e.: *more than usual*; *same as usual*; *less than usual*; and *much*

less than usual) and the NW items (i.e.: *not at all*; *no more than usual*; *rather more than usual*; and *much more than usual*). Accordingly, the 4–point scoring scheme was applied in our study so total scores in the GHQ–12 ranged from 0 to a maximum of 36, with higher scores indicating lower levels of GPH.

Table 1. *Descriptive statistics, standardized factor loadings from Model 6 and correlations between the Model 6 factors and the covariates*

			Model 6		
Item	*M*	*SD*	*GPH*	*PW*	*NW*
Item 1. Able to concentrate	1.03	0.37	.49*	.50*	
Item 2. Lost sleep over worry	0.57	0.75	.78*		.13
Item 3. Playing a useful part in things	0.96	0.31	.09*	.47*	
Item 4. Capable of making decisions	0.96	0.30	.22*	.76*	
Item 5. Constantly under strain	0.71	0.79	.85*		.06
Item 6. Could not overcome difficulties	0.44	0.66	.71*		.31*
Item 7. Enjoy day–to–day activities	1.01	0.40	.56*	.57*	
Item 8. Face up to problems	0.99	0.32	.39*	.62*	
Item 9. Feeling unhappy and depressed	0.37	0.66	.76*		.43*
Item 10. Losing confidence in yourself	0.19	0.48	.52*		.83*
Item 11. Thinking of yourself as a worthless person	0.12	0.40	.49*		.72*
Item 12. Feeling reasonably happy	0.99	0.38	.42*	.58*	
Correlations between the Model 6 factors and the socio-demographic variables					
Sex			.18*	-.14*	.02
Age			.13*	.04	.03
Educational level			-.04	-.13*	-.12*

Note. *M* = Mean; *SD* = Standard Deviation; *GPH* = General Health Psychology factor; *PW* = Positive Wording factor; *NW* = Negative Wording factor. *$p<.05$

For the purposes of exploring the correlates of method effects (i.e., item wording effects), we used the following three covariates: (a) sex; (b) age; and (c) educational level, which was measured as a self–reported question with 7 response graduated categories ranging from *incomplete primary studies* to *postgraduate studies*. The educational level was scored as the highest level of education reached.

Statistical Analysis

A set of competing confirmatory factor models were estimated using LISREL 8.70[55]. Figure 1 shows the specification of all these CFA models. Models 1 and 2 are the best fitted in Molina et al.[36], so they are included here as base-models for the purpose of comparison. These two models examine the method effect associated with NW items: Model 1 is a unidimensional model with correlated errors (i.e., a CTCU model); and Model 2 is a unidimensional model with an additional factor for the NW items (i.e., a CTCM model). Analogously, two models were estimated as a means of examining the method effects associated with PW items (Models 3 and 4, respectively). Finally, Model 5 and Model 6 take into account the method effects associated with both NW and PW items: the former as a CTCU model, the latter as a CTCM model.

The goodness–of–fit indices computed were: the chi–square statistic; the Comparative Fit Index (CFI); the Tucker–Lewis Index (TLI); the root mean square error of approximation (RMSEA) with its 90% confidence interval; the standardized root mean square residual (SRMR); and the Akaike information criterion (AIC). Values greater than 0.95 for CFI and TLI, and lower than 0.06 and 0.08 for RMSEA and SRMR, respectively, are considered to indicate good model fit. The model with the lowest AIC is considered to be the best one. Moreover, the chi–square difference test was computed to decide between competing nested models.

As concerns the estimation of CFA models, most studies into the GHQ–12 factor structure have used maximum likelihood[16,31,35,40,44]. This estimation method relies on several

assumptions which should be met to be confident about the results obtained. This is the case of the assumption of multivariate normality which implies, first, that the variables are continuous in nature and, second, that the joint distribution of the variables is normal. The first condition is unlikely to be met with the GHQ–12 Likert–type response data; nor is the second if the variables depart markedly from normality as is the case for the responses to the NW items which were heavily positively skewed (see Figure 2). An alternative when these conditions are not met is to use the weighted least squares (WLS) estimator[56], which has already been used in some studies about the GHQ–12 factor structure[13,18,20,29] and it will be the estimation method used here. Thus, the various CFA models were estimated using WLS, after computing the respective polychoric correlation and asymptotic covariance matrices.

Finally, correlates of the GHQ–12 factors were evaluated using SEM through the inclusion in the finally selected model of the 3 covariates considered in this study: sex was treated as categorical, whereas age and educational level were treated as continuous variables.

Figure 2

Results

The goodness–of–fit statistics obtained for the 6 models compared here are shown in Table 2.

Table 2. *Fit indexes for the alternative models of the 12–item General Health Questionnaire*

Models	*df*	Chi-square	CFI	RMSEA [90% CI]	TLI	SRMR	AIC
Model 1	39	226.96	.99	.041 [.036, .046]	.97	.082	304.96
Model 2	48	458.86	.97	.054 [.050, .059]	.96	.095	518.86
Model 3	39	371.93	.97	.054 [.049, .059]	.96	.120	449.93
Model 4	48	435.32	.97	.052 [.048, .057]	.96	.140	495.32
Model 6	41	152.41	.99	.030 [.025, .036]	.99	.072	226.41

Note. Models are specified in Figure 1. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; TLI = Tucker–Lewis Index; SRMR = Standardized Root Mean Square Residual: AIC = Akaike Information Criteria

As for the comparison of the models that include a latent method factor for the NW or the PW items (Models 2 and 4, respectively), it may be observed that both models demonstrated a similar fit according to all the goodness–of–fit indices, except for the SRMR that was lower for Model 5, but above the cut point of 0.08. However, when comparing the corresponding CTCU models (Models 1 and 3), the model that includes the correlated uniquenesses among the NW items (Model 1) fitted better than the model that includes the correlated uniquenesses among the PW items (Model 36) according to all the fit indices –significantly better if we compare their respective RMSEA 90% ICs. Model 5 did not converge to a fully proper solution, thereby making it impossible either to compare Model 5 with its nested models (Models 1 and 3), or to compare it with Model 6. Finally, Model 6, which includes two method factors for the PW and the NW items, was the model which provided the best fit according to all the fit indices. When comparing Model 6 with the nested Models 2 and 4, the chi–square difference test was statistically significant for both comparisons (306.45 (7); $p < .001$ for Models 2 and 6; and 282.91 (7); $p < .001$ for Models 4 and 6). An in–depth inspection of the parameter estimates in Model 6 (see Table 1) showed that all factor loadings were statistically significant for the three factors, except for items 2 and 5 in the method factor comprising the NW items. The correlation between the two method factors was also statistically significant ($r = .20$).

Finally, a statistical analysis of the relationships between the latent factors in Model 6 and the 3 covariates considered in this study (i.e. sex, age and educational level) was performed through a SEM in which the correlations between the 3 latent factors in Model 6 and the 3 covariates were freely estimated, the focus being on the relationships between the method factors and the covariates. The model fit was good (RMSEA = .034; RMSEA 90% IC = [.030, .038]; CFI = .98; NNFI = .97; SRMR = .079). As can be seen in Table 1, the correlations of age with the method factors were near to 0 and statistically non–significant. However, sex was significantly correlated with the method

factor associated with PW items (–.14), whereas the educational level was significantly related to both method factors (–.12 and –.13). Thus, men and women differ in the way they answer PW items, meaning that men are more likely than women to endorse PW items. Additionally, subjects with higher educational levels are less prone to showing a bias associated with the wording of items, regardless of whether the items are positively or negatively worded.

Discussion

This study focused on the examination of the latent structure underlying the responses to the GHQ–12, considering the role of method effects associated with both, positive and negative items wording, and using two alternative parameterizations of the CFA measurement models. What should first be noted is that the studies that have included method effects in the measurement model of the GHQ–12 have been more the exception than the rule in previous research into the factor structure of this questionnaire.

According to the results of the present study, we conclude that the GHQ–12 factor structure is best characterized by introducing latent method factors that capture both the method effects associated with NW and PW items (Model 6). Moreover, the statistically significant correlation between the two method factors (r = .20) suggests, as was the case in the work by Van Dam et al.[57] on the Five Facet Mindfulness Questionnaire, that respondents susceptible to negative method effects are also susceptible to positive method effects. These results support the conclusion from previous research that the good fit obtained by multidimensional models (mainly the two–factor model and the three–factor Graetz's model,) could simply be explained by the artificial grouping of PW and NW items.

The second aim of this study was to examine the relationships between the method factors associated with both NW and PW items and three demographic variables, namely, the sex, age and educational level of the respondents. Regarding the sex, we found a statistically significant, but

weak, relationship between PW and sex, so that men were more likely than women to endorse PW items. These results are in line with previous works that, in the context of RSES, have found sex differences in wording effects[49,50]. As for the explanatory role of age on method effects, we found that the relationship between age and the negative wording effect was not statistically significant, which supports previous research using other questionnaires (e.g., self–esteem scales,[58]; Hospital Anxiety & Depression Scale,[59]). However, our results give no support to previous studies which had stated that, in older adults, the strongest method effects would be associated with PW items, rather than NW items[48,51].

As to the educational level, we found that there was a significant correlation of this variable on the two method factors, so that less educated participants were more prone to showing a bias associated with the wording of the items, regardless of whether the items were positively or negatively worded, while more educated respondents would treat negatively and positively worded items more equally. This result supports and extends the evidence obtained in previous research on the relationship of the negative wording factor and the educational level/verbal ability with different questionnaires and samples[41,59–64]. The size of the effects found in this work (–.13 and –.12) were similar to those found in Wouters et al.[59] for the NW items (–.12 to –.15 for models with a different number of trait factors). Contrary to the above results, Tomás et al.[58] found that the educational level of the respondents had no effect on the negative method factor using self–esteem questionnaires; however, as they pointed out, the indicator used was a coarse measure of this variable and more research would be desirable, making use of more reliable indicators.

Taken together, the results on the individual differences related to the demographic variables considered in this study cannot only help to understand the presence of wording method effects but also to identify respondents who are prone to answering PW and NW items differently. In this sense, the relationship that appears as more evident is for the educational level. The relationship that appears as more evident in previous works with other questionnaires and in the present work, is for

the educational level, suggesting that the effect of educational level on the responses to NW and PW items could be invariant instead of questionnaire-specific; further research is needed in this regard.

Another practical consequence of our study concerns the relationship between the intended measure of the GHQ–12 (i.e., the GPH factor) and other constructs of interest. Several studies have shown that method effects can inflate, deflate or have no effect at all on estimates of the relationship between two constructs (see Podsakoff[65], for a further review of the effects that method biases have on individual measures and on the covariation between different constructs). Thus, it is desirable that both the constructs of interest as well as the effects of method factors, like positive and negative wording, are considered in SEM models as a means of controlling these systematic sources of bias and, thus, avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

As was the case in Hankins[31], it is interesting to note that we found more extreme scores for NW than for PW items, so that the overall mean for NW items was lower (0.40) than that for PW items (0.99). Thus, it follows that the respondents assessed their psychological health more positively when answering NW items than when answering PW items. This asymmetry in the participants' responses as a function of the wording of the items is consistent with results from previous research into wording effects for contrastive survey questions[66]. The extent to which the presence of method effects is linked to this asymmetric pattern of responses to PW and NW items in the GHQ–12 should be examined in future research.

Comparing the current work with previous studies into the factorial structure of the GHQ–12, to our knowledge, this is the first study that, on the one hand, tests a comprehensive set of models including method effects associated with both PW and NW items and also explores some demographic correlates of these method effects. Another strength of this work was the fact that it used a large representative sample of workers, but the results might not be generalized to other

specific populations, for example, adolescents and elderly retired people.

Funding

This work was supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

No conflict of interest has been declared by the authors.

Contributors

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data. GM and MFR: drafted the article. JV and J-ML: critically revised the draft for important intellectual content. All authors agreed on the final version.

References

1 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: : Oxford University Press 1972.

2 Goldberg DP, Williams P. *A user's guide to the General Health Questionnaire*. Windsor, United Kingdom: : NFER-Nelson 1988.

3 Bhui K, Bhugra D, Goldberg D. Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Soc Psychiatry Psychiatr Epidemiol* 2000;**35**:248–54.

4 Daradkeh TK, Ghubash R, El-Rufaie O. Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychol Rep* 2001;**89**:85–94.

5 Gelaye B, Tadesse MG, Lohsoonthorn V, *et al.* Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J Affect Disord* 2015;**187**:197–202. doi:10.1016/J.JAD.2015.08.045

6 Goldberg DP, Gater R, Sartorius N, *et al.* The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;**27**:191–7.

7 Lundin A, Hallgren M, Theobald H, *et al.* Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health* 2016;**136**:66–74. doi:10.1016/J.PUHE.2016.03.005

8 Rocha K, Pérez K, Rodríguez-Sanz M, *et al.* Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *Int J Clin Heal Psychol* 2011;**11**:125–39.

9 Sanchez-Lopez MP, Dresch V. The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema* 2008;**20**:839–43.

10 Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire

(GHQ-12) in a German primary care sample. *Acta Psychiatr Scand* 1999;**100**:462–8.

11 Tait RJ, French DJ, Hulse GK. Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents. *Aust N Z J Psychiatry* 2003;**37**:374–81.

12 Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol* 1991;**26**:132–8.

13 Campbell A, Knowles S. A confirmatory factor analysis of the GHQ12 using a large Australian sample. *Eur J Psychol Assess* 2007;**23**:2–8.

14 Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *Int J Geriatr Psychiatry* 2002;**17**:739–44.

15 French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004;**13**:1–7.

16 Gao F, Luo N, Thumboo J, *et al.* Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health Qual Life Outcomes* 2004;**2**:1–7. doi:10.1186/1477-7525-2-63

17 Mäkikangas A, Feldt T, Kinnunen U, *et al.* The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. *Psychol Assess* 2006;**18**:444–51.

18 Padrón A, Galán I, Durbán M, *et al.* Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Qual Life Res* 2012;**21**:1291–8.

19 Penninkilampi-Kerola V, Miettunen J, Ebeling H. Health and disability: A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scand J Psychol* 2006;**47**:431–40.

20 Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychol Assess* 2005;**17**:231–6.

21 Martin CR, Newell RJ. The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement. *J Psychosom Res* 2005;**59**:193–9. doi:10.1016/j.jpsychores.2005.02.020

22 Tomás JM, Meléndez JC, Oliver A, *et al.* Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores. *Psicológica* 2010;**31**:383–400.

23 Andrich D, Van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychol Med* 1989;**19**:469–85.

24 Gao W, Stark D, Bennett MI, *et al.* Using the 12-item General Health Questionnaire to screen psychological distress from survivorship to end-of-life care: Dimensionality and item quality. *Psychooncology* 2012;**21**:954–61.

25 Glozah FN, Pevalin DJ. Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *J Child Adolesc Ment Heal* 2015;**27**:53–7. doi:10.2989/17280583.2015.1007867

26 Kilic C, Rezaki M, Rezaki B, *et al.* General Health Questionnaire (GHQ-12 & GHQ-28): Psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol* 1997;**32**:327.

27 Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: Reliability, validity and factor structure of the GHQ-12. *Eur Acad Dermatology Venereol* 2001;**15**:410.

28 Werneke U, Goldberg DP, Yalcin I, *et al.* The stability of the factor structure of the General Health Questionnaire. *Psychol Med* 2000;**30**:823.

29 Aguado J, Campbell A, Ascaso C, *et al.* Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish post-partum women. *Assessment* 2012;**19**:517–25.

30 Ye S. Factor structure of the General Health Questionnaire (GHQ-12): The role of wording

effects. *Pers Individ Dif* 2009;**46**:197–201.

31 Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clin Pract Epidemiol Ment Heal* 2008;**4**:10.

32 Harvey RJ, Billings RS, Nilan KJ. Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *J Appl Psychol* 1985;**70**:461–8.

33 Smith N, Stults DM. Factors defined by negatively keyed items: The results of careless respondents? *Appl Psychol Meas* 1985;**9**:367–73.

34 Abubakar A, Fischer R. The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress Heal J Int Soc Investig Stress* 2012;**28**:248–54.

35 Fernandes HM, Vasconcelos-Raposo J. Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment* 2013;**20**:219–29. doi:http://dx.doi.org/10.1177/1073191112465768

36 Molina JG, Rodrigo MF, Losilla J-M, *et al.* Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychol Assess* 2014;**26**:1031–1037. doi:10.1037/a0036472

37 Motamed N, Edalatian Zakeri S, Rabiee B, *et al.* The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Appl Res Qual Life* 2018;**13**:303–16. doi:10.1007/s11482-017-9522-y

38 Rey JJ, Abad FJ, Barrada JR, *et al.* The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychol Assess* 2014;**26**:1021–30. doi:10.1037/a0036468

39 Romppel M, Braehler E, Roth M, *et al.* What is the General Health Questionnaire-12 assessing? *Compr Psychiatry* 2013;**54**:406–13. doi:10.1016/j.comppsych.2012.10.010

40 Smith AB, Oluboyede Y, West R, *et al.* The factor structure of the GHQ-12: The interaction between item phrasing, variance and levels of distress. *Qual Life Res An Int J Qual Life Asp Treat Care Rehabil* 2013;**22**:145–52.

41 Marsh HW. Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Dev Psychol* 1986;**22**:37–49.

42 Tomás JM, Oliver A. Rosenberg's Self-Esteem Scale: Two factors or method effects. *Struct Equ Model A Multidiscip J* 1999;**6**:84–98.

43 Gnambs T, Staufenbiel T. The structure of the General Health Questionnaire (GHQ-12): two meta-analytic factor analyses. *Health Psychol Rev* 2018;**12**:179–94. doi:10.1080/17437199.2018.1426484

44 Wang L, Lin W. Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Pers Individ Dif* 2011;**50**:1056–61.

45 Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychol Assess* 2010;**22**:366–81.

46 DiStefano C, Motl RW. Self-Esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Struct Equ Model A Multidiscip J* 2009;**16**:134–46.

47 Gana K, Saada Y, Bailly N, *et al.* Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *J Res Pers* 2013;**47**:406–16.

48 Lindwall M, Barkoukis V, Grano C, *et al.* Method effects: The problem with negatively versus positively keyed items. *J Pers Assess* 2012;**94**:196–204.

49 Michaelides MP, Zenger M, Koutsogiorgi C, *et al.* Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Pers Individ Dif* 2016;**97**:13–8. doi:10.1016/J.PAID.2016.03.011

50 Urbán R, Szigeti R, Kökönyei G, *et al.* Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behav Res*

Methods 2014;**46**:488–98. doi:10.3758/s13428-013-0391-5

51 Mullen SP, Gothe NP, McAuley E. Evaluation of the factor structure of the Rosenberg Self-Esteem Scale in older adults. *Pers Individ Dif* 2013;**54**:153–7.

52 Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 1996;**70**:810–9. doi:10.1037/0022-3514.70.4.810

53 Catalonian Labor Relations and Quality of Work Department. *Segunda Encuesta Catalana de Condiciones de Trabajo [Second Catalonian Survey of Working Conditions]*. Barcelona: : Autor 2012.

54 Lobo A, Muñoz PE. Versiones en lengua española validadas. In: Goldberg D, Williams P, eds. *Cuestionario de Salud General GHQ (General Health Questionnaire)*. Guia para el usuario de las distintas versiones [Guide for the use of the different validated versions in Spanish language]. Barcelona, Spain: : Editorial Masson 1996.

55 Jöreskog K, Sörbom D. *Lisrel 8.70*. Chicago: : Scientific Software International Inc 2004.

56 Browne MW. Asymptotic distribution free methods in the analysis of covariance structures. *Br J Math Stat Psychol* 1984;**37**:62–83.

57 Van Dam NT, Hobkirk AL, Danoff-Burg S, *et al.* Mind your words: Positive and negative items create method effects on the Five Facet Mindfulness Questionnaire. *Assessment* 2012;**19**:198–204. doi:http://dx.doi.org/10.1177/1073191112438743

58 Tomás JM, Oliver A, Galiana L, *et al.* Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Struct Equ Model A Multidiscip J* 2013;**20**:299–313.

59 Wouters E, Booysen F le R, Ponnet K, *et al.* Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PLoS One* 2012;**7**:1–7. doi:10.1371/journal.pone.0034881

60 Bors DA, Vigneau FO, Lalande F. Measuring the Need for Cognition: Item polarity,

dimensionality, and the relation with ability. *Pers Individ Dif* 2006;**40**:819–28.

61 Chen YH, Rendina-Gobioff G, Dedrick RF. Factorial Invariance of a Chinese Self-Esteem Scale for Third and Sixth Grade Students: Evaluating Method Effects Associated with Positively and Negatively Worded Items. *Int J Educ Psychol Assess* 2010;**6**:21–35.

62 Corwyn RF. The factor structure of global self-esteem among adolescents and adults. *J Res Pers* 2000;**34**:357–79.

63 Rammstedt B, Goldberg LR, Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *J Res Pers* 2010;**44**:53–61.

64 Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *J Pers Soc Psychol* 2005;**89**:623–42.

65 Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 2012;**63**:539–69.

66 Kamoen N, Holleman B, Mak P, *et al.* Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions. *Discourse Process* 2011;**48**:355–85. doi:10.1080/0163853X.2011.578910

→FILE Figure1.pdf

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

→FILE Figure2.pdf

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

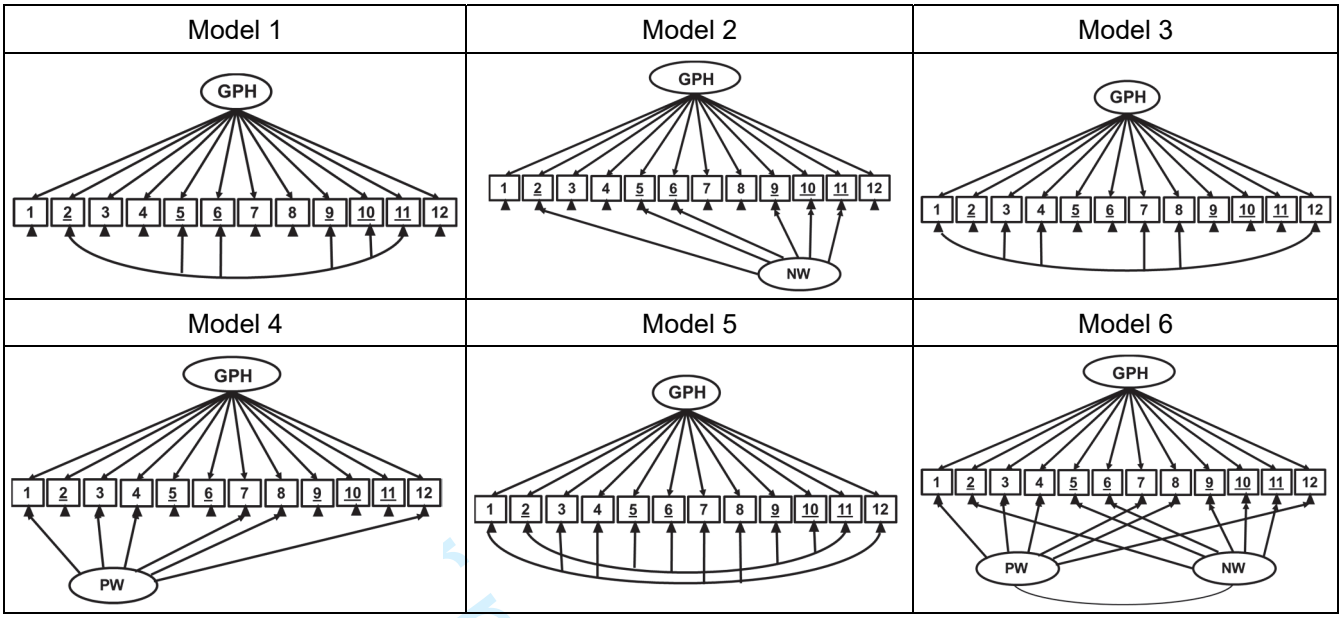

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

Positively worded items

Negatively worded items

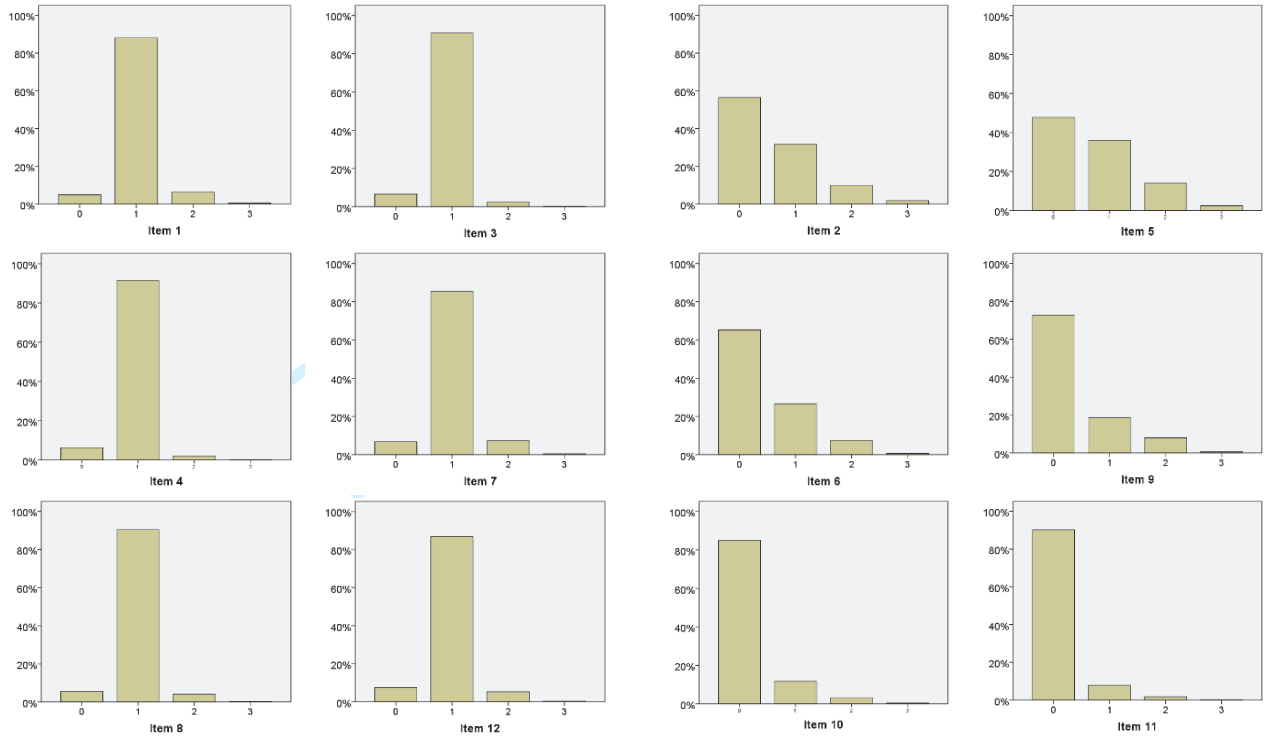

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

BMJ Open

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative sample of Catalonian workers

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031859.R1
Article Type:	Original research
Date Submitted by the Author:	30-Jul-2019
Complete List of Authors:	Rodrigo, Maria F.; University of Valencia Molina, J. Gabriel; University of Valencia Losilla, Josep-Maria; Autonomous University of Barcelona Vives, Jaume; Autonomous University of Barcelona Tomás, José; University of Valencia
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Research methods, Mental health
Keywords:	psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

1. **Title**

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative sample of Catalonian workers

2. Authors and affiliation

Rodrigo, Maria F.
Department of Methodology for the Behavioral Sciences
University of Valencia
Valencia (Spain)

Molina, J. Gabriel †
Department of Methodology for the Behavioral Sciences
University of Valencia
Valencia (Spain)

Losilla, Josep-Maria
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Barcelona (Spain)

Vives, Jaume
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Barcelona (Spain)

Tomás, José M.
Department of Methodology for the Behavioral Sciences
University of Valencia
Valencia (Spain)

†Deceased September 26, 2014

3. Word count, excluding title page, abstract, references, figures and tables: 3119

4. Authors' statements

All authors have agreed to authorship in the indicated order.
All authors declare that this paper is an original unpublished work and it is not being submitted elsewhere.
All authors do not have any financial interests that might be interpreted as influencing the research, and APA ethical standard were followed in the conduct of the study.

The research was not submitted to approval by an institutional review board since this is not a requirement at our universities for this type of study.

5. Data availability statement

Data is publicly available upon request to the Catalonian Labor Relations and Quality of Work Department.

6. Contact information for the corresponding author:

Jaume Vives
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Edifici B - Despatx B5b/081
Carrer de la Fortuna / Carrer de Ca n'Altayó. Campus de la UAB
08193 Bellaterra (Cerdanyola del Vallès). Barcelona (Spain)
E-mail: jaume.vives@uab.cat
Phone: +34 93 581 23 31

Abstract

Objective. Recent studies into the factorial structure of the 12–item version of the General Health Questionnaire (GHQ–12) have shown that it was best represented by a single substantive factor when method effects associated with negatively worded (NW) items are considered. The purpose of the present study was to examine the presence of method effects, and their relationships with demographic covariates, associated with positively worded (PW) and/or NW items.

Method. The current work compared a comprehensive set of confirmatory factor models, including method effects associated with PW and/or NW items with GHQ-12 responses using a random sample of 3050 workers.

Results. A confirmatory factor analysis showed that the best–fitting model was a unidimensional model with two additional uncorrelated method factors associated with PW and NW items. Furthermore, structural equation modeling revealed that method effects were differentially related to both the sex and age of the respondents.

Conclusion. Individual differences related to sex and age can help to identify respondents who are prone to answering PW and NW items differently. Consequently, is desirable that both the constructs of interest as well as the effects of method factors are considered in SEM models as a means of avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

Keywords

psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

Strengths and limitations of this study

Strengths

- Sampling quality: A random and large representative sample of workers and face-to-face administration by professional interviewers.
- To compare confirmatory models for positively and/or negatively worded items and using two different parameterizations.
- To study demographic correlates of wording effects.

Limitations

- The different response scale used for the NW items and the PW items in the questionnaire could be a confusion variable.
- The results might not be generalized to other specific populations as, for example, adolescents and elderly retired people.

Introduction

Originally developed by Goldberg[1], the General Health Questionnaire (GHQ) has been widely used as a screening instrument for measuring General Psychological Health (GPH) in both community and non–psychiatric clinical settings[2]. The shortest 12–item version (GHQ–12) is the most popular and has been employed on different settings and in several countries, as well as part of multiple major national health, social wellbeing and occupational surveys, achieving results which underline the fact that it is highly reliable and valid[3–11].

Despite its broad application, the factor structure underlying the responses to the GHQ–12 remains a controversial issue. In this sense, although the GHQ–12 was originally developed as a unidimensional scale, this one–factor latent structure has found little empirical support and some alternative multidimensional models have been proposed as more appropriate. Thus, the one with the most empirical support is the three–factor model proposed by Graetz[5,12–22]. It is important to note that the 6 positively worded (PW) items make up the first factor, whereas the other two factors are made up of the 6 negatively worded (NW) items (see Figure 1, Model 8). On the other hand, the bidimensional model, where the 6 NW and the 6 PW items in the GHQ–12 are grouped into two factors, has also obtained wide support, especially in studies based on exploratory factor analysis[5,10,23–28]. The arguments against these models and in favor of the unidimensional solution are the high correlations between the factors[13] and the low discriminant validity of the factor scores derived from these models[16,29,30].

As Hankins[31] points out, multifactor models may just be the resulting artifact of the inclusion of PW and NW items in the questionnaire and so, the controversy about the factorial structure of the GHQ–12 might relate to the effect of item wording on subjects' response patterns as part of a more general category called 'method[32,33]. Hankins[31] found that, after modeling the wording effects for the NW items, the unidimensional model fitted better than both the two–factor model (NW vs. PW items) and Graetz's three–factor model. Other studies have called into question

the substantive meaning of the GHQ–12 multifactor solutions, suggesting that they might just be an artifact due to the wording effects associated with NW items[29,30,34–40]. See Molina et al.[36] for a deeper review about the dimensionality of GHQ-12.

Some studies about other instruments, however, suggested not only considering the wording effects for the NW items but also for the PW items[41,42]. Regarding GHQ-12, only a recent analytical factor meta-analysis modelled the presence of method effects for negatively and positively worded items concluding that positively keyed items explained incremental variance beyond a general mental health factor[43].

Another source of variability in the results about the factor structure of the GHQ-12 could come from the statistical control of method biases, that has been mainly achieved through the correlated traits–correlated methods (CTCM) and the correlated traits–correlated uniquenesses (CTCU) confirmatory factor analysis models. Both procedures have been used in GHQ–12, to deal with method effects applying the CTCM model[i.e., 30,44] the CTCU model[i.e., 29,31,39,40], or both CTCM and CTCU[i.e., 34–37].

To date, we have not found any studies about GHQ-12 that analyze the wording effects associated with either PW items alone, nor with NW and PW items simultaneously, comparing both CTCU and CTCM models. There are several multivariate statistical models for analyzing method effect, and among them the CFA based approaches are the most popular approach[45]. Among the different CFA models stand out the CFA with correlated traits and correlated methods (CFA-CTCM) and the CFA with correlated traits and correlated uniqueness (CTCU). The CTCM specifies that item (indicator) variance can be explained by can be written as a linear combination of trait, method, and error effects[46]. Trait and method effects are treated as latent variables. The CTCM model, when method result uncorrelated (or are specified as independent) translates into the well-known Bifactor model[47,48]. The CTCU model specifies trait factors while method effects are modeled correlating the uniqueness of items (indicators) sharing a common method[49]. Both

CTCM and CTCU models have strengths and shortcomings and therefore are usually employed simultaneously[50]. So, this work extends the previous work by Molina et al.[36], which compares the fit of the unidimensional model, the multifactor models and the CTCM and CTCU unidimensional models with method effects for only the NW items.

To clarify this work, Figure 1 (Model 1 to Model 9) shows the 9 CFA models that we consider here in order to test the potential method effects associated with either the PW or the NW or both. Model 1 is a one factor model of general health. This model also works as a baseline model against which to compare other more complex models. Models 2 and 3 are the CTCU and CTCM models that include method effects for the NW items. These were the best fitting models in Molina et al. [36] and are the base-models for this study. Models 4 and 5 are the CTCU and CTCM models including method effects for the PW items. Model 6 is the CTCM model including method factors for both the NW and PW items (a CTCU model with method effects for both PW and NW items was not estimated because it is not identified). Model 7 is a bifactor model with a general trait factor of general health and two method factors associated to NW and PW items. The three factors are independent (uncorrelated). Additionally, and considering the best fitting multidimensional model in Tomás, Gutiérrez and Sancho[51] based on the results by Graetz [12], models 8 and 9 were also tested. Model 8 posited three substantive dimensions: social dysfunction, anxiety and depression and loss of confidence. Model 9 included an additional method factor associated to NW items. Models considering a method factor associated to PW items made no sense as all PW items were indicators of social dysfunction.

As stressed by Marsh et al.[52], it becomes necessary to consider this comprehensive set of competing models to determine the relative importance and substantive nature of the method effects.

Figure 1

Finally, there has been some research carried out on the demographic correlates of method effects, such as sex[53–57], age[55,58] or educational level[41,59]. With respect to the GHQ–12, to date, we have not found any studies that analyze demographic correlates of method effects.

Building on the previous studies, the first aim of this study was to overcome the limitation pointed out in Molina et al.[36] and examine method effects associated with both positive and negative wording. The second aim was to further understand the meaning of the method factors; therefore, we evaluated the relationships between the method factors and three covariates (i.e., the sex, age, and educational level) in the framework of a structural equation model (SEM).

Method

Participants

The data used in this study came from the Second Catalonian Survey of Working Conditions[60] and were based on a representative random sample of all employees living in Catalonia (Spain). Data were collected between September and November 2010 by professional interviewers in private households. The sample comprised a total of 3,050 participants who responded to the GHQ–12 included in the survey. Main sociodemographic characteristics of the sample are shown in Table 1.

Table 1. *Main sociodemographic characteristics*.

	M (SD)	n (%)	Range
Gender			
Women		1361 (44.6)	
Age	40.46 (11.19)		17-82
Education			
Incomplete primary studies		90 (3.0)	
Primary studies		541 (17.9)	
Secondary studies. 1st stage		637 (21.0)	
Associate degree		763 (25.2)	

High School	598 (19.8
Graduate studies	359 (11.9)
Postgraduate studies	39 (1.3)

Note. *M* = Mean; *SD* = Standard Deviation

Public Involvement

Respondents were not involved in any stage of the design of the study and were only requested to respond the survey. In the selected households, interviewers identified themselves personally and informed this was an official survey about the working conditions of employed Catalonian people commissioned by the Catalonian Government Work Department.

Results were published on the Catalonian Government Work Department website[60] and are available at https://treball.gencat.cat/ca/ambits/seguretat_i_salut_laboral/publicacions/estadistiques_estudis/ci/ii_ecct/treballadors/

Measures

The GHQ–12 is a self–report scale that contains 6 PW items (e.g. "Have you been able to face up to problems?") and 6 NW items (e.g. "Have you been losing confidence in yourself?"). The GHQ–12 was validated in Spain by Lobo and Muñoz[61]. Table 2 shows the statements of these items in the same order as they were presented in the survey. It must be noted that the GHQ–12 has a different response scale for the PW items (i.e.: *more than usual*; *same as usual*; *less than usual*; and *much less than usual*) and the NW items (i.e.: *not at all*; *no more than usual*; *rather more than usual*; and *much more than usual*). Accordingly, the 4–point scoring scheme was applied in our study so total scores in the GHQ–12 ranged from 0 to a maximum of 36, with higher scores indicating lower levels of GPH.

Table 2. *Descriptive statistics, standardized factor loadings from Model 7 and correlations between the Model 7 factors and the covariates*

			Model 7		
Item	*M*	*SD*	*GPH*	*PW*	*NW*

Item 1. Able to concentrate	1.03	0.37	.42*	.49*	
Item 2. Lost sleep over worry	0.57	0.75	.78*		.07
Item 3. Playing a useful part in things	0.96	0.31	.09*	.59*	
Item 4. Capable of making decisions	0.96	0.30	.14*	.70*	
Item 5. Constantly under strain	0.71	0.79	.83*		.03
Item 6. Could not overcome difficulties	0.44	0.66	.76*		.25*
Item 7. Enjoy day–to–day activities	1.01	0.40	.53*	.55*	
Item 8. Face up to problems	0.99	0.32	.39*	.60*	
Item 9. Feeling unhappy and depressed	0.37	0.66	.78*		.38*
Item 10. Losing confidence in yourself	0.19	0.48	.53*		.70*
Item 11. Thinking of yourself as a worthless person	0.12	0.40	.48*		.72*
Item 12. Feeling reasonably happy	0.99	0.38	.44*	.72*	

Relations between the Model 7 factors and the socio-demographic variables

Sex	.13*	-.08*	-.02
Age	.11*	.08*	.01
Educational level	.00	-.02	-.06

Note. *M* = Mean; *SD* = Standard Deviation; *GPH* = General Health Psychology factor; *PW* = Positive Wording factor; *NW* = Negative Wording factor. *$p < .05$

For the purposes of exploring the correlates of method effects (i.e., item wording effects), we used the following three covariates: (a) sex; (b) age; and (c) educational level, which was measured as a self–reported question with 7 response graduated categories ranging from *incomplete primary studies* to *postgraduate studies*. The educational level was scored as the highest level of education reached.

Statistical Analysis

A set of competing confirmatory factor models were estimated using MPlus 8.3[62]. Figure 1 shows the specification of all these CFA models. The goodness–of–fit indices computed were: the chi–square statistic; the Comparative Fit Index (CFI); the Root Mean Square Error of Approximation (RMSEA) with its 90% confidence interval; and the Standardized Root Mean Square Residual (SRMR). Values greater than 0.95 for CFI and TLI, and lower than 0.06 and 0.08 for RMSEA and SRMR, respectively, are considered to indicate good model fit.

As concerns the estimation of CFA models, most studies into the GHQ–12 factor structure have used maximum likelihood[16,31,35,40,44]. This estimation method relies on several assumptions which should be met to be confident about the results obtained. This is the case of the assumption of multivariate normality which implies, first, that the variables are continuous in nature and, second, that the joint distribution of the variables is normal. The first condition is unlikely to be met with the GHQ–12 Likert–type response data; nor is the second if the variables depart markedly from normality as is the case for the responses to the NW items which were heavily positively skewed (see Figure 2). An alternative when these conditions are not met is to use the weighted least squares (WLS) estimator[63], which has already been used in some studies about the GHQ–12 factor structure[13,18,20,29] and it will be the estimation method used here. Thus, the various CFA models were estimated using Diagonally WLS.

Finally, correlates of the GHQ–12 factors were evaluated using SEM through the inclusion

in the finally selected model of the 3 covariates considered in this study: sex was treated as categorical, whereas age and educational level were treated as continuous variables.

Figure 2

Results

The goodness–of–fit statistics and indices obtained for the 9 models compared here are shown in Table 3.

Table 3. *Fit indexes for the alternative models of the 12–item General Health Questionnaire*

Models	*df*	Chi-	CFI	RMSEA [90% CI]	SRMR
Model 1	54	5378.68	.77	.180 [.176, .184]	.119
Model 2	39	928.099	.96	.086 [.082, .091]	.049
Model 3	48	1345.38	.95	.094 [.090, .059]	.061
Model 4	39	934.690	.96	.087 [.083, .092]	.052
Model 5	48	1275.28	.95	.092 [.087, .096]	.058
Model 6	41	497.520	.98	.060 [.056, .065]	.030
Model 7	42	507.741	.98	.060 [.056, .065]	.030
Model 8	51	1142.88	.95	.084 [.080, .088]	.054
Model 9	45	960.388	.96	.082 [.078, .086]	.049

Notes: Models are specified in Figure 1; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual.

Model 1, with a single factor of general health, and model 8, with three substantive factors, had worse fit than the models that include wording effects. That is, a careful look at fit indexes makes clear that the inclusion of method effects always improves model fit. Indeed, both NW and PW method effects are needed to get the best fitting models. These best fitting models were models 6 and 7. Their fit was practically indistinguishable and, given that they only differ in that model 7 is more parsimonious because constrains method factors correlation to zero, it will be retained as the best representation of the observed data.

An in–depth inspection of the parameter estimates in Model 7 (see Table 2) showed that all factor loadings were statistically significant for the three factors, except for items 2 and 5 in the method factor comprising the NW items.

Finally, a statistical analysis of the relationships between the latent factors in Model 7 and the 3 covariates considered in this study (i.e. sex, age and educational level) was performed through a MIMIC SEM model in which the effects between the 3 latent factors in Model 7 and the 3 covariates were freely estimated, the focus being on the relationships between the method factors and the covariates. The model fit was excellent (RMSEA = .040; RMSEA 90% IC = [.037, .049]; CFI = .99; SRMR = .029). As can be seen in Table 2, the relations of age with the method factors were near to 0 and statistically non–significant for NW items, and positive and significant although small with PW items (.08). Sex was significantly related with the method factor associated with PW items (–.08), whereas the educational level was not significantly related to method factors. Thus, men and women differ in the way they answer PW items, meaning that men are slightly more likely than women to endorse PW items, and method effects associated to PW items also increased by age.

Discussion

This study focused on the examination of the latent structure underlying the responses to the GHQ–12, considering the role of method effects associated with both, positive and negative items wording, and using two alternative parameterizations of the CFA measurement models. What should first be noted is that the studies that have included method effects in the measurement model of the GHQ–12 have been more the exception than the rule in previous research into the factor structure of this questionnaire.

According to the results of the present study, we conclude that the GHQ–12 factor structure is best characterized by introducing latent method factors that capture both the method effects associated with NW and PW items (Model 7). These results support the conclusion from previous

research that the good fit obtained by multidimensional models (mainly the two–factor model and the three–factor Graetz's model,) could simply be explained by the artificial grouping of PW and NW items.

The second aim of this study was to examine the relationships between the method factors associated with both NW and PW items and three demographic variables, namely, the sex, age and educational level of the respondents. Regarding the sex, we found a statistically significant, but weak, relationship between PW and sex, so that men were more likely than women to endorse PW items. These results are in line with previous works that, in the context of RSES, have found sex differences in wording effects[56,57]. As for the explanatory role of age on method effects, we found that the relationship between age and the negative wording effect was not statistically significant, which supports previous research using other questionnaires (e.g., self–esteem scales,[50]; Hospital Anxiety & Depression Scale,[64]). Moreover, our results give support to previous studies which had stated that, in older adults, the strongest method effects would be associated with PW items, rather than NW items[55,58].

As to the educational level, we found that there was not a significant correlation of this variable on the two method factors. This result supports and extends the evidence obtained in Tomás et al.[50], that found that the educational level of the respondents had no effect on the negative method factor using self–esteem questionnaires. This results contradicts previous research on the relationship of the negative wording factor and the educational level/verbal ability with different questionnaires and samples[41,64–69].

Taken together, the results on the individual differences related to the demographic variables considered in this study cannot only help to understand the presence of wording method effects but also to identify respondents who are prone to answering PW and NW items differently. In this sense, the relationship that appears as more evident is for the age and sex variables.

Another practical consequence of our study concerns the relationship between the intended

measure of the GHQ–12 (i.e., the GPH factor) and other constructs of interest. Several studies have shown that method effects can inflate, deflate or have no effect at all on estimates of the relationship between two constructs (see Podsakoff[70], for a further review of the effects that method biases have on individual measures and on the covariation between different constructs). Thus, it is desirable that both the constructs of interest as well as the effects of method factors, like positive and negative wording, are considered in SEM models as a means of controlling these systematic sources of bias and, thus, avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

Previous research on the GHQ-12 (e.g.[31,36]) has outlined the asymmetry in the participants' responses as a function of the wording of the items, as well as the different responses scales for the positive and negative items. This asymmetry in the participants' responses as a function of the wording of the items is consistent with results from previous research into wording effects for contrastive survey questions[71]. The extent to which the presence of method effects is linked to the asymmetric pattern of responses and/or to the different response scales for the PW and NW items in the GHQ–12 should be examined in future research.

Comparing the current work with previous studies into the factorial structure of the GHQ–12, to our knowledge, this is the first study that, on the one hand, tests a comprehensive set of models including method effects associated with both PW and NW items and also explores some demographic correlates of these method effects. Another strength of this work was the fact that it used a large representative sample of workers, but the results might not be generalized to other specific populations, for example, adolescents and elderly retired people.

Funding

This work was supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

No conflict of interest has been declared by the authors.

Contributors

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data. MFR and GM: drafted the article. JV and JML: critically revised the draft for important intellectual content. JMT worked in the statistical analysis and interpretation of data. All authors agreed on the final version.

References

1 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: : Oxford University Press 1972.

2 Goldberg DP, Williams P. *A user's guide to the General Health Questionnaire*. Windsor, United Kingdom: : NFER-Nelson 1988.

3 Bhui K, Bhugra D, Goldberg D. Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Soc Psychiatry Psychiatr Epidemiol* 2000;**35**:248–54.

4 Daradkeh TK, Ghubash R, El-Rufaie O. Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychol Rep* 2001;**89**:85–94.

5 Gelaye B, Tadesse MG, Lohsoonthorn V, *et al.* Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J Affect Disord* 2015;**187**:197–202. doi:10.1016/J.JAD.2015.08.045

6 Goldberg DP, Gater R, Sartorius N, *et al.* The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;**27**:191–7.

7 Lundin A, Hallgren M, Theobald H, *et al.* Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health* 2016;**136**:66–74. doi:10.1016/J.PUHE.2016.03.005

8 Rocha K, Pérez K, Rodríguez-Sanz M, *et al.* Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *Int J Clin Heal Psychol* 2011;**11**:125–39.

9 Sanchez-Lopez MP, Dresch V. The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema* 2008;**20**:839–43.

10 Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatr Scand* 1999;**100**:462–8.

11 Tait RJ, French DJ, Hulse GK. Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents. *Aust N Z J Psychiatry* 2003;**37**:374–81.

12 Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol* 1991;**26**:132–8.

13 Campbell A, Knowles S. A confirmatory factor analysis of the GHQ12 using a large Australian sample. In: *European Journal of Psychological Assessment*. 2007. 2–8.

14 Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *Int J Geriatr Psychiatry* 2002;**17**:739–44.

15 French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004;**13**:1–7.

16 Gao F, Luo N, Thumboo J, *et al.* Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health Qual Life Outcomes* 2004;**2**:1–7. doi:10.1186/1477-7525-2-63

17 Mäkikangas A, Feldt T, Kinnunen U, *et al.* The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. *Psychol Assess* 2006;**18**:444–51.

18 Padrón A, Galán I, Durbán M, *et al.* Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Qual Life Res* 2012;**21**:1291–8.

19 Penninkilampi-Kerola V, Miettunen J, Ebeling H. Health and disability: A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scand J Psychol* 2006;**47**:431–40.

20 Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: A

large sample analysis using confirmatory factor analysis. *Psychol Assess* 2005;**17**:231–6.

21 Martin CR, Newell RJ. The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement. *J Psychosom Res* 2005;**59**:193–9. doi:10.1016/j.jpsychores.2005.02.020

22 Tomás JM, Meléndez JC, Oliver A, *et al.* Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores. *Psicológica* 2010;**31**:383–400.

23 Andrich D, Van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychol Med* 1989;**19**:469–85.

24 Gao W, Stark D, Bennett MI, *et al.* Using the 12-item General Health Questionnaire to screen psychological distress from survivorship to end-of-life care: Dimensionality and item quality. *Psychooncology* 2012;**21**:954–61.

25 Glozah FN, Pevalin DJ. Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *J Child Adolesc Ment Heal* 2015;**27**:53–7. doi:10.2989/17280583.2015.1007867

26 Kilic C, Rezaki M, Rezaki B, *et al.* General Health Questionnaire (GHQ-12 & GHQ-28): Psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol* 1997;**32**:327.

27 Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: Reliability, validity and factor structure of the GHQ-12. *Eur Acad Dermatology Venereol* 2001;**15**:410.

28 Werneke U, Goldberg DP, Yalcin I, *et al.* The stability of the factor structure of the General Health Questionnaire. *Psychol Med* 2000;**30**:823.

29 Aguado J, Campbell A, Ascaso C, *et al.* Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish post-partum women. *Assessment* 2012;**19**:517–25.

30 Ye S. Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Pers Individ Dif* 2009;**46**:197–201.

31 Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clin Pract Epidemiol Ment Heal* 2008;**4**:10.

32 Harvey RJ, Billings RS, Nilan KJ. Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *J Appl Psychol* 1985;**70**:461–8.

33 Smith N, Stults DM. Factors defined by negatively keyed items: The results of careless respondents? *Appl Psychol Meas* 1985;**9**:367–73.

34 Abubakar A, Fischer R. The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress Heal J Int Soc Investig Stress* 2012;**28**:248–54.

35 Fernandes HM, Vasconcelos-Raposo J. Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment* 2013;**20**:219–29. doi:http://dx.doi.org/10.1177/1073191112465768

36 Molina JG, Rodrigo MF, Losilla J-M, *et al.* Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychol Assess* 2014;**26**:1031–1037. doi:10.1037/a0036472

37 Motamed N, Edalatian Zakeri S, Rabiee B, *et al.* The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Appl Res Qual Life* 2018;**13**:303–16. doi:10.1007/s11482-017-9522-y

38 Rey JJ, Abad FJ, Barrada JR, *et al.* The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychol Assess* 2014;**26**:1021–30. doi:10.1037/a0036468

39 Romppel M, Braehler E, Roth M, *et al.* What is the General Health Questionnaire-12 assessing? *Compr Psychiatry* 2013;**54**:406–13. doi:10.1016/j.comppsych.2012.10.010

40 Smith AB, Oluboyede Y, West R, *et al.* The factor structure of the GHQ-12: The interaction between item phrasing, variance and levels of distress. *Qual Life Res An Int J Qual Life Asp*

Treat Care Rehabil 2013;**22**:145–52.

41 Marsh HW. Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Dev Psychol* 1986;**22**:37–49.

42 Tomás JM, Oliver A. Rosenberg's Self-Esteem Scale: Two factors or method effects. *Struct Equ Model A Multidiscip J* 1999;**6**:84–98.

43 Gnambs T, Staufenbiel T. The structure of the General Health Questionnaire (GHQ-12): two meta-analytic factor analyses. *Health Psychol Rev* 2018;**12**:179–94. doi:10.1080/17437199.2018.1426484

44 Wang L, Lin W. Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Pers Individ Dif* 2011;**50**:1056–61.

45 Wothke W. Models for multitrait-multimethod matrix analysis. In: Marcoulides GA, Schumacker RE, eds. *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, NJ: : Lawrence Erlbaum Associates 1996.

46 Jöreskog KG. Analyzing psychological data by structural analysis of covariance matrices. In: Atkinson RC, Krantz DH, Luce RD, *et al.*, eds. *Contemporary developments in mathematical psychology. Vol. 2*. San Francisco: : Freeman 1974. 1–56.

47 Markon KE. Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol* 2019;**15**:51–69. doi:10.1146/annurev-clinpsy-050718-095522

48 Reise SP. The Rediscovery of Bifactor Measurement Models. *Multivariate Behav Res* 2012;**47**:667–96. doi:10.1080/00273171.2012.715555

49 Marsh HW, Bailey M. Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models. *Appl Psychol Meas* 1991;**15**:47–70. doi:10.1177/014662169101500106

50 Tomás JM, Oliver A, Galiana L, *et al.* Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Struct Equ*

Model A Multidiscip J 2013;**20**:299–313.

51 Tomás JM, Gutiérrez M, Sancho P. Factorial Validity of the General Health Questionnaire 12 in an Angolan Sample. *Eur J Psychol Assess* 2017;**33**:116–22. doi:10.1027/1015-5759/a000278

52 Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychol Assess* 2010;**22**:366–81.

53 DiStefano C, Motl RW. Self-Esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Struct Equ Model A Multidiscip J* 2009;**16**:134–46.

54 Gana K, Saada Y, Bailly N, *et al.* Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *J Res Pers* 2013;**47**:406–16.

55 Lindwall M, Barkoukis V, Grano C, *et al.* Method effects: The problem with negatively versus positively keyed items. *J Pers Assess* 2012;**94**:196–204.

56 Michaelides MP, Zenger M, Koutsogiorgi C, *et al.* Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Pers Individ Dif* 2016;**97**:13–8. doi:10.1016/J.PAID.2016.03.011

57 Urbán R, Szigeti R, Kökönyei G, *et al.* Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behav Res Methods* 2014;**46**:488–98. doi:10.3758/s13428-013-0391-5

58 Mullen SP, Gothe NP, McAuley E. Evaluation of the factor structure of the Rosenberg Self-Esteem Scale in older adults. *Pers Individ Dif* 2013;**54**:153–7.

59 Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 1996;**70**:810–9. doi:10.1037/0022-3514.70.4.810

60 Catalonian Labor Relations and Quality of Work Department. *Segunda Encuesta Catalana de Condiciones de Trabajo [Second Catalonian Survey of Working Conditions]*. Barcelona: : Author 2012.

61 Lobo A, Muñoz PE. Versiones en lengua española validadas. In: Goldberg D, Williams P, eds. *Cuestionario de Salud General GHQ (General Health Questionnaire)*. Guia para el usuario de las distintas versiones [Guide for the use of the different validated versions in Spanish language]. Barcelona, Spain: : Editorial Masson 1996.

62 Muthén LK, Muthén BO. *Mplus User's Guide*. 6th ed. Los Angeles, CA: : Muthén & Muthén 2011.

63 Browne MW. Asymptotic distribution free methods in the analysis of covariance structures. *Br J Math Stat Psychol* 1984;**37**:62–83.

64 Wouters E, Booysen F le R, Ponnet K, *et al.* Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PLoS One* 2012;**7**:1–7. doi:10.1371/journal.pone.0034881

65 Bors DA, Vigneau FO, Lalande F. Measuring the Need for Cognition: Item polarity, dimensionality, and the relation with ability. *Pers Individ Dif* 2006;**40**:819–28.

66 Chen YH, Rendina-Gobioff G, Dedrick RF. Factorial Invariance of a Chinese Self-Esteem Scale for Third and Sixth Grade Students: Evaluating Method Effects Associated with Positively and Negatively Worded Items. *Int J Educ Psychol Assess* 2010;**6**:21–35.

67 Corwyn RF. The factor structure of global self-esteem among adolescents and adults. *J Res Pers* 2000;**34**:357–79.

68 Rammstedt B, Goldberg LR, Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *J Res Pers* 2010;**44**:53–61.

69 Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *J Pers*

Soc Psychol 2005;**89**:623–42.

70 Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 2012;**63**:539–69.

71 Kamoen N, Holleman B, Mak P, *et al.* Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions. *Discourse Process* 2011;**48**:355–85. doi:10.1080/0163853X.2011.578910

→FILE Figure1.pdf

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

→FILE Figure2.pdf

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

Positively worded items Negatively worded items

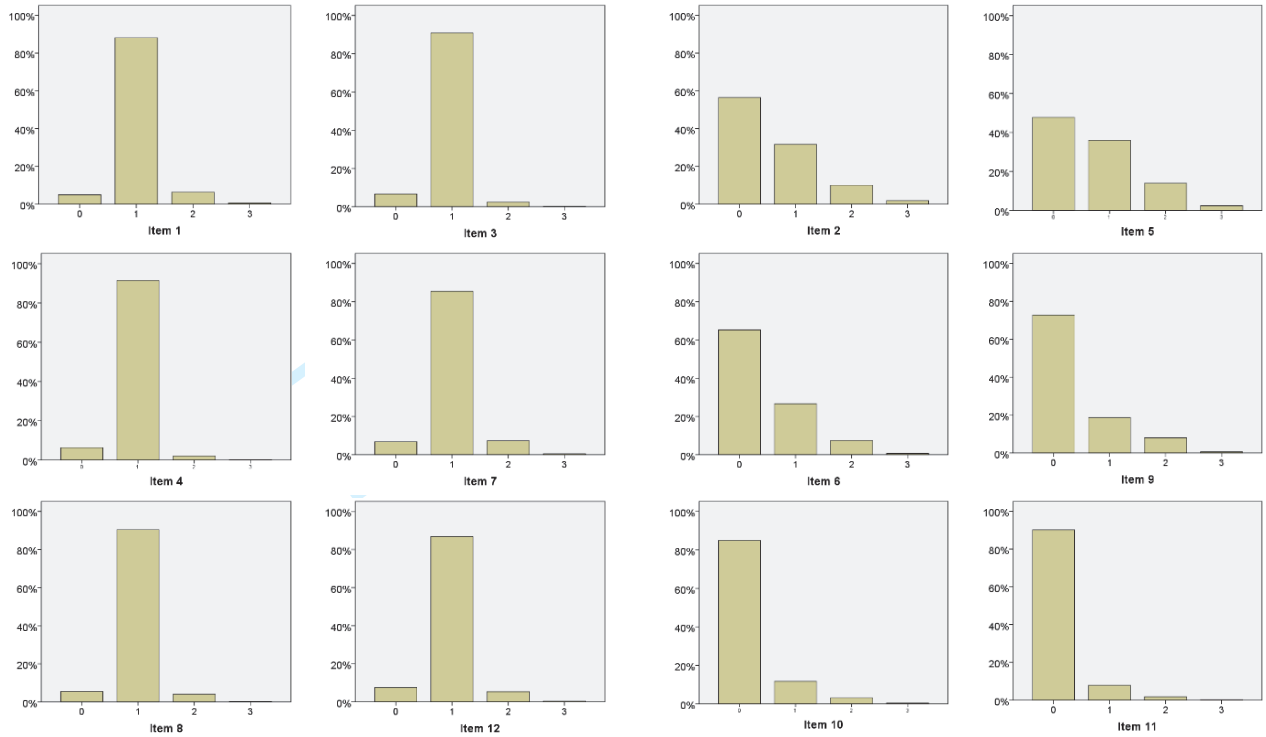

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

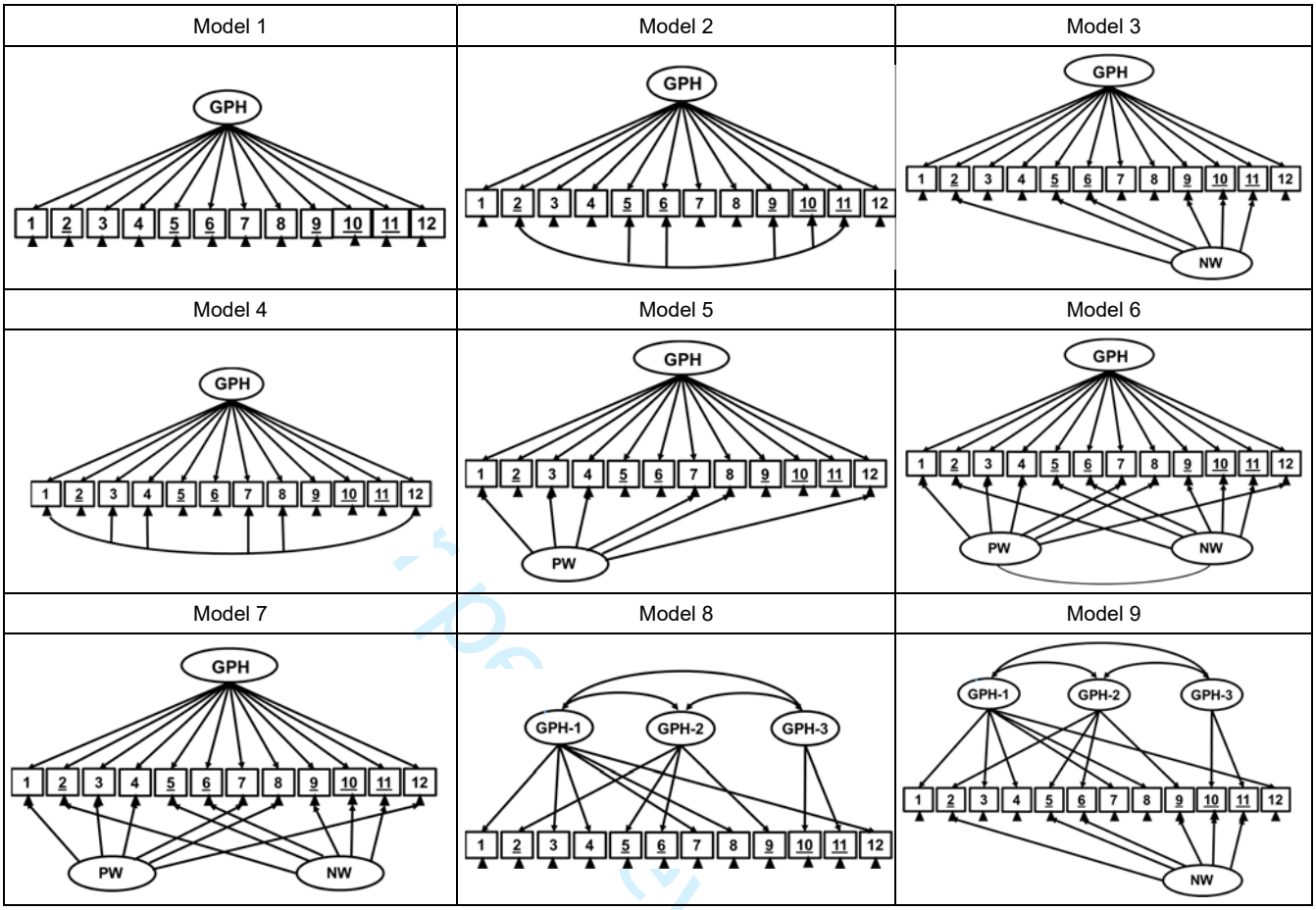

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; GPH–1: Social dysfunction; GPH–2: Anxiety and depression; GPH–3: Loss of confidence; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract P.1 Lines 1-4 (Title)
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found P.3 Lines 1-18 (Abstract)
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported P.5 - P.8 (Introduction)
Objectives	3	State specific objectives, including any prespecified hypotheses P.8 Lines 5-9 (Introduction)
Methods		
Study design	4	Present key elements of study design early in the paper P.8 Lines 13-14 (Method - Participants)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection P.8 Lines 15-16 (Method - Participants)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants P.8 Lines 14-15 (Method - Participants)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable P.9 Lines 2-11 and P.11 Lines 1-5 (Method - Measures)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group P.8 Lines 14-16 and P.9 Lines 2-11 (Method - Measures)
Bias	9	Describe any efforts to address potential sources of bias P.8 Line 14 (Method - Participants)
Study size	10	Explain how the study size was arrived at P.8 Line 13 (Method - Participants)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why P.11 Lines 14-24 and P.12 Line 2 (Method – Statistical Analysis)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding P.11 Lines 7-24 (Method – Statistical Analysis)
		(*b*) Describe any methods used to examine subgroups and interactions P.11 Line 25 and P.12 Lines 1-2 (Method – Statistical Analysis)
		(*c*) Explain how missing data were addressed P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy P.8 Line 13 (Method – Participants)
		(*e*) Describe any sensitivity analyses P.11 Lines 14-24 (Method – Statistical Analysis)
Results		

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed P.8 Lines 16-18 (Method – Participants)
		(b) Give reasons for non-participation at each stage P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
		(c) Consider use of a flow diagram P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders P.8 Table 1 (Method – Participants)
		(b) Indicate number of participants with missing data for each variable of interest P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
Outcome data	15*	Report numbers of outcome events or summary measures P.9 Table 2 (Method - Measures) and P.12 Figure 2 (Method – Statistical Analysis)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included P.9 Table 2 (Method - Measures) and P.12 Table 3 (Results)
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses P.13 Lines 4-14 (Results)
Discussion		
Key results	18	Summarise key results with reference to study objectives P.13 Lines 23-25 and P.14 Lines 1-14 (Discussion)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias P.4 Lines 10-14 (Limitations) and P.15 Lines 20-21 (Discussion)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence P.14 Lines 21-25 and P.15 Lines 1-15 (Discussion)
Generalisability	21	Discuss the generalisability (external validity) of the study results P.15 Lines 16-20 (Discussion)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based P.16 Lines 1-5 (Funding)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative sample of Catalonian workers

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031859.R2
Article Type:	Original research
Date Submitted by the Author:	03-Sep-2019
Complete List of Authors:	Rodrigo, Maria F.; University of Valencia Molina, J. Gabriel; University of Valencia Losilla, Josep-Maria; Autonomous University of Barcelona Vives, Jaume; Autonomous University of Barcelona Tomás, José; University of Valencia
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Research methods, Mental health
Keywords:	psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

1 **Title**

2 Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General
3 Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative
4 sample of Catalonian workers
5

6 2. Authors and affiliation

7 Rodrigo, Maria F.
8 Department of Methodology for the Behavioral Sciences
9 University of Valencia
10 Valencia (Spain)
11
12 Molina, J. Gabriel †
13 Department of Methodology for the Behavioral Sciences
14 University of Valencia
15 Valencia (Spain)
16
17 Losilla, Josep-Maria
18 Department of Psychobiology and Methodology of Health Sciences
19 Universitat Autònoma de Barcelona
20 Barcelona (Spain)
21
22 Vives, Jaume
23 Department of Psychobiology and Methodology of Health Sciences
24 Universitat Autònoma de Barcelona
25 Barcelona (Spain)
26
27 Tomás, José M.
28 Department of Methodology for the Behavioral Sciences
29 University of Valencia
30 Valencia (Spain)
31
32 †Deceased September 26, 2014
33

34 3. Word count, excluding title page, abstract, references, figures and tables: 3119
35

36 4. Authors' statements

37 All authors have agreed to authorship in the indicated order.
38 All authors declare that this paper is an original unpublished work and it is not being submitted
39 elsewhere.
40 All authors do not have any financial interests that might be interpreted as influencing the research,
41 and APA ethical standard were followed in the conduct of the study.
42

5. Ethics statement.
The research was not submitted to approval by an institutional review board since this is not a requirement at our universities for this type of study.
Ethics approval was not sought for this study since this was a secondary analysis of anonymized data.

6. Data availability statement
Data is publicly available upon request to the Catalonian Labor Relations and Quality of Work Department.

7. Contact information for the corresponding author:

Jaume Vives
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Edifici B - Despatx B5b/081
Carrer de la Fortuna / Carrer de Ca n'Altayó. Campus de la UAB
08193 Bellaterra (Cerdanyola del Vallès). Barcelona (Spain)
E-mail: jaume.vives@uab.cat
Phone: +34 93 581 23 31

Abstract

Objective. Recent studies into the factorial structure of the 12–item version of the General Health Questionnaire (GHQ–12) have shown that it was best represented by a single substantive factor when method effects associated with negatively worded (NW) items are considered. The purpose of the present study was to examine the presence of method effects, and their relationships with demographic covariates, associated with positively worded (PW) and/or NW items.

Design. A cross-sectional, observational study to compare a comprehensive set of confirmatory factor models, including method effects associated with PW and/or NW items with GHQ-12 responses.

Setting. Representative sample of all employees living in Catalonia (Spain).

Participants. 3050 participants (44.6% women) who responded the Second Catalonian Survey of Working Conditions.

Results. A confirmatory factor analysis showed that the best–fitting model was a unidimensional model with two additional uncorrelated method factors associated with PW and NW items. Furthermore, structural equation modeling revealed that method effects were differentially related to both the sex and age of the respondents.

Conclusion. Individual differences related to sex and age can help to identify respondents who are prone to answering PW and NW items differently. Consequently, is desirable that both the constructs of interest as well as the effects of method factors are considered in SEM models as a means of avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

Keywords

psychological health, General Health Questionnaire (GHQ–12), method effects, item wording

effects, confirmatory factor analysis

Strengths and limitations of this study

Strengths

- Sampling quality: A random and large representative sample of workers and face-to-face administration by professional interviewers.
- Comparison of confirmatory models for positively and/or negatively worded items and the use of two different parameterizations. Previous works on the GHQ-12 have not included such a set of competing models.
- Investigation of demographic correlates of wording effects. There are no previous works on this subject on the GHQ-12.

Limitations

- The different response scale used for the NW items and the PW items in the questionnaire could be a confusion variable.
- The results might not be generalized to other specific populations as, for example, adolescents and elderly retired people.

Introduction

Originally developed by Goldberg[1], the General Health Questionnaire (GHQ) has been widely used as a screening instrument for measuring General Psychological Health (GPH) in both community and non–psychiatric clinical settings[2]. The shortest 12–item version (GHQ–12) is the most popular and has been employed on different settings and in several countries, as well as part of multiple major national health, social wellbeing and occupational surveys, achieving results which underline the fact that it is highly reliable and valid[3–11].

Despite its broad application, the factor structure underlying the responses to the GHQ–12 remains a controversial issue. In this sense, although the GHQ–12 was originally developed as a unidimensional scale, this one–factor latent structure has found little empirical support and some alternative multidimensional models have been proposed as more appropriate. Thus, the one with the most empirical support is the three–factor model proposed by Graetz[5,12–22]. It is important to note that the 6 positively worded (PW) items make up the first factor, whereas the other two factors are made up of the 6 negatively worded (NW) items (see Figure 1, Model 8). On the other hand, the bidimensional model, where the 6 NW and the 6 PW items in the GHQ–12 are grouped into two factors, has also obtained wide support, especially in studies based on exploratory factor analysis [5,10,23–28]. The arguments against these models and in favor of the unidimensional solution are the high correlations between the factors [13] and the low discriminant validity of the factor scores derived from these models[16,29,30].

As Hankins[31] points out, multifactor models may just be the resulting artifact of the inclusion of PW and NW items in the questionnaire and so, the controversy about the factorial structure of the GHQ–12 might relate to the effect of item wording on subjects' response patterns as part of a more general category called 'method[32,33]. Hankins[31] found that, after modeling the wording effects for the NW items, the unidimensional model fitted better than both the two–factor model (NW vs. PW items) and Graetz's three–factor model. Other studies have called into question

the substantive meaning of the GHQ–12 multifactor solutions, suggesting that they might just be an artifact due to the wording effects associated with NW items[29,30,34–40]. See Molina et al.[36] for a deeper review about the dimensionality of GHQ-12.

Some studies about other instruments, however, suggested not only considering the wording effects for the NW items but also for the PW items[41,42]. Regarding GHQ-12, only a recent meta-analysis modelled the presence of method effects for negatively and positively worded items concluding that positively keyed items explained incremental variance beyond a general mental health factor[43].

Therefore, another source of variability in the results about the factor structure of the GHQ-12 could come from the statistical control of method biases, that has been mainly achieved through the correlated traits–correlated methods (CTCM) and the correlated traits–correlated uniquenesses (CTCU) confirmatory factor analysis models. Both procedures have been used in GHQ–12, to deal with method effects applying the CTCM model [i.e., 30,44] the CTCU model [i.e., 29,31,39,40], or both CTCM and CTCU [i.e., 34–37].

To date, we have not found any study about GHQ-12 that analyze the wording effects associated with either PW items alone, nor with NW and PW items simultaneously, comparing both CTCU and CTCM models. There are several multivariate statistical models for analyzing method effect, and among them the CFA based approaches are the most popular ones [45]. Among the different CFA models stand out the CFA with correlated traits and correlated methods (CFA-CTCM) and the CFA with correlated traits and correlated uniqueness (CTCU). On one hand, the CTCM model specifies that indicators' variance can be explained by a linear combination of trait, method, and error effects [46], with trait and method effects specified as latent variables. The CTCM model, when methods are specified independent (uncorrelated), directly translates into the well-known Bifactor model [47,48]. On the other hand, the CTCU model specifies trait factors while method effects are modeled correlating the uniqueness of items (indicators) sharing a

common method [49]. Both CTCM and CTCU models have strengths and shortcomings and therefore are usually employed simultaneously [50]. This work extends the previous work by Molina et al. [36], which compares the fit of the unidimensional model, the multifactor models and the CTCM and CTCU unidimensional models with method effects for only the NW items.

To clarify, Figure 1 (Model 1 to Model 9) shows the 9 CFA models estimated to test the potential method effects associated with either the PW or the NW, or both. Model 1 is a one factor model of general health. This model also works as a baseline model against which to compare other more complex models. Models 2 and 3 are the CTCU and CTCM models that include method effects for the NW items. These were the best fitting models in Molina et al. [36]. Models 4 and 5 are the CTCU and CTCM models including method effects for the PW items. Model 6 is the CTCM model including method factors for both the NW and PW items (a CTCU model with method effects for both PW and NW items was not estimated because it is not identified). Model 7 is a bifactor model with a general trait factor of general health and two method factors associated to NW and PW items. The three factors are independent (uncorrelated). Additionally, and considering the best fitting multidimensional model in Tomás, Gutiérrez and Sancho [51] based on the results by Graetz [12], models 8 and 9 were also tested. Model 8 posited three substantive dimensions: social dysfunction, anxiety and depression, and loss of confidence. Model 9 included an additional method factor associated to NW items. Models considering a method factor associated to PW items made no sense as all PW items were indicators of social dysfunction.

As stressed by Marsh et al. [52], it becomes necessary to consider this comprehensive set of competing models to determine the relative importance and substantive nature of the method effects.

Figure 1

Finally, there has been some research carried out on the demographic correlates of method

effects, such as sex[53–57], age[55,58] or educational level[41,59]. With respect to the GHQ–12, to date, we have not found any studies that analyze demographic correlates of method effects.

Building on the previous studies, the first aim of this study was to overcome the limitation pointed out in Molina et al.[36] and examine method effects associated with both positive and negative wording. The second aim was to further understand the meaning of the method factors; therefore, we evaluated the relationships between the method factors and three covariates (i.e., the sex, age, and educational level) in the framework of a structural equation model (SEM).

Method

Participants

The data used in this study came from the Second Catalonian Survey of Working Conditions[60] and were based on a representative random sample of all employees living in Catalonia (Spain). Data were collected between September and November 2010 by professional interviewers in private households. The sample comprised a total of 3,050 participants who responded to the GHQ–12 included in the survey. Main sociodemographic characteristics of the sample are shown in Table 1.

Table 1. *Main sociodemographic characteristics*.

	M (SD)	n (%)	Range
Gender			
Women		1361 (44.6)	
Age	40.46 (11.19)		17-82
Education			
Incomplete primary studies		90 (3.0)	
Primary studies		541 (17.9)	
Secondary studies. 1st stage		637 (21.0)	
Associate degree		763 (25.2)	
High School		598 (19.8	
Graduate studies		359 (11.9)	
Postgraduate studies		39 (1.3)	

Note. *M* = Mean; *SD* = Standard Deviation

Public Involvement

Respondents were not involved in any stage of the design of the study and were only requested to respond the survey. In the selected households, interviewers identified themselves personally and informed this was an official survey about the working conditions of employed Catalonian people commissioned by the Catalonian Government Work Department.

Results were published on the Catalonian Government Work Department website[60] and are available at https://treball.gencat.cat/ca/ambits/seguretat_i_salut_laboral/publicacions/estadistiques_estudis/ci/ii_ecct/treballadors/

Measures

The GHQ–12 is a self–report scale that contains 6 PW items (e.g. "Have you been able to face up to problems?") and 6 NW items (e.g. "Have you been losing confidence in yourself?"). The GHQ–12 was validated in Spain by Lobo and Muñoz[61]. Table 2 shows the statements of these items in the same order as they were presented in the survey. It must be noted that the GHQ–12 has a different response scale for the PW items (i.e.: *more than usual*; *same as usual*; *less than usual*; and *much less than usual*) and the NW items (i.e.: *not at all*; *no more than usual*; *rather more than usual*; and *much more than usual*). Accordingly, the 4–point scoring scheme was applied in our study so total scores in the GHQ–12 ranged from 0 to a maximum of 36, with higher scores indicating lower levels of GPH.

Table 2. *Descriptive statistics, standardized factor loadings from Model 7 and correlations between the Model 7 factors and the covariates*

			Model 7		
Item	*M*	*SD*	*GPH*	*PW*	*NW*
Item 1. Able to concentrate	1.03	0.37	.42*	.49*	
Item 2. Lost sleep over worry	0.57	0.75	.78*		.07

Item 3. Playing a useful part in things	0.96	0.31	.09*	.59*	
Item 4. Capable of making decisions	0.96	0.30	.14*	.70*	
Item 5. Constantly under strain	0.71	0.79	.83*		.03
Item 6. Could not overcome difficulties	0.44	0.66	.76*		.25*
Item 7. Enjoy day–to–day activities	1.01	0.40	.53*	.55*	
Item 8. Face up to problems	0.99	0.32	.39*	.60*	
Item 9. Feeling unhappy and depressed	0.37	0.66	.78*		.38*
Item 10. Losing confidence in yourself	0.19	0.48	.53*		.70*
Item 11. Thinking of yourself as a worthless person	0.12	0.40	.48*		.72*
Item 12. Feeling reasonably happy	0.99	0.38	.44*	.72*	

Relations between the Model 7 factors and the socio-demographic variables

Sex	.13*	-.08*	-.02
Age	.11*	.08*	.01
Educational level	.00	-.02	-.06

Note. *M* = Mean; *SD* = Standard Deviation; *GPH* = General Health Psychology factor; *PW* = Positive Wording factor; *NW* = Negative Wording factor. *$p < .05$

For the purposes of exploring the correlates of method effects (i.e., item wording effects), we used the following three covariates: (a) sex (0= men and 1= women); (b) age; and (c) educational level, which was measured as a self–reported question with 7 response graduated categories ranging from *incomplete primary studies* to *postgraduate studies*. The educational level was scored as the highest level of education reached.

Statistical Analysis

A set of competing confirmatory factor models were estimated using MPlus 8.3[62]. Figure 1 shows the specification of all these CFA models. The goodness–of–fit indices computed were: the chi–square statistic; the Comparative Fit Index (CFI); the Root Mean Square Error of Approximation (RMSEA) with its 90% confidence interval; and the Standardized Root Mean Square Residual (SRMR). Values greater than 0.95 for CFI and TLI, and lower than 0.06 and 0.08 for RMSEA and SRMR, respectively, are considered to indicate good model fit.

As concerns the estimation of CFA models, most studies into the GHQ–12 factor structure have used maximum likelihood[16,31,35,40,44]. This estimation method relies on several assumptions which should be met to be confident about the results obtained. This is the case of the assumption of multivariate normality which implies, first, that the variables are continuous in nature and, second, that the joint distribution of the variables is normal. The first condition is unlikely to be met with the GHQ–12 Likert–type response data; nor is the second if the variables depart markedly from normality as is the case for the responses to the NW items which were heavily positively skewed (see Figure 2). An alternative when these conditions are not met is to use the weighted least squares (WLS) estimator [63], which has already been used in some studies about the GHQ–12 factor structure[13,18,20,29] and it will be the estimation method used here. Thus, the various CFA models were estimated using Diagonally WLS.

Finally, correlates of the GHQ–12 factors were evaluated using SEM through the inclusion

in the finally selected model of the 3 covariates considered in this study: sex was treated as categorical, whereas age and educational level were treated as continuous variables.

Figure 2

Results

The goodness–of–fit statistics and indices obtained for the 9 models compared here are shown in Table 3.

Table 3. *Fit indexes for the alternative models of the 12–item General Health Questionnaire*

Models	*df*	Chi-	CFI	RMSEA [90% CI]	SRMR
Model 1	54	5378.68	.77	.180 [.176, .184]	.119
Model 2	39	928.099	.96	.086 [.082, .091]	.049
Model 3	48	1345.38	.95	.094 [.090, .059]	.061
Model 4	39	934.690	.96	.087 [.083, .092]	.052
Model 5	48	1275.28	.95	.092 [.087, .096]	.058
Model 6	41	497.520	.98	.060 [.056, .065]	.030
Model 7	42	507.741	.98	.060 [.056, .065]	.030
Model 8	51	1142.88	.95	.084 [.080, .088]	.054
Model 9	45	960.388	.96	.082 [.078, .086]	.049

Notes: Models are specified in Figure 1; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual.

Model 1, with a single factor of general health, and model 8, with three substantive factors, had worse fit than the models that include wording effects. That is, a careful look at fit indexes makes clear that the inclusion of method effects always improves model fit. Indeed, both NW and PW method effects are needed to get the best fitting models. These best fitting models were models 6 and 7. Their fit was practically indistinguishable and, given that they only differ in that model 7 is more parsimonious because constrains method factors correlation to zero, it will be retained as the best representation of the observed data.

An in–depth inspection of the parameter estimates in Model 7 (see Table 2) showed that all factor loadings were statistically significant for the three factors, except for items 2 and 5 in the method factor comprising the NW items.

Finally, a statistical analysis of the relationships between the latent factors in Model 7 and the 3 covariates considered in this study (i.e. sex, age and educational level) was performed through a MIMIC SEM model in which the effects between the 3 latent factors in Model 7 and the 3 covariates were freely estimated, the focus being on the relationships between the method factors and the covariates. The model fit was excellent (RMSEA = .040; RMSEA 90% IC = [.037, .049]; CFI = .99; SRMR = .029). As can be seen in Table 2, the relations of age with the method factors were near to 0 and statistically non–significant for NW items, and positive and significant although small with PW items (.08). Sex was significantly related with the method factor associated with PW items (–.08), whereas the educational level was not significantly related to method factors. Thus, men and women differ in the way they answer PW items, meaning that men are slightly more likely than women to endorse PW items, and method effects associated to PW items also increased by age.

Discussion

This study focused on the examination of the latent structure underlying the responses to the GHQ–12, considering the role of method effects associated with both, positive and negative items wording, and using two alternative parameterizations of the CFA measurement models. What should first be noted is that the studies that have included method effects in the measurement model of the GHQ–12 have been more the exception than the rule in previous research into the factor structure of this questionnaire.

According to the results of the present study, we conclude that the GHQ–12 factor structure is best characterized by introducing latent method factors that capture both the method effects associated with NW and PW items (Model 7). These results support the conclusion from previous

research that the good fit obtained by multidimensional models (mainly the two–factor model and the three–factor Graetz's model) could simply be explained by the artificial grouping of PW and NW items. However, the interpretation of the latent (method) factors as purely integrating method bias due to wording is not straightforward. It is obvious that NW and PW items share the wording. It is also clear that this three bifactor model (one trait and two method factors) fitted the data best. And finally, there is a lot of empirical evidence on these wording effects. However, it is also relevant to discuss the large loadings of many items on the method factors, being these loadings sometimes larger than their loadings in the trait factor. The general factor explains a 52% of the shared variance, but there are some items that deserve careful attention. For example, items 3 ("playing useful part in things") and 4 ("capable of making decisions") had very low loads on the trait factor. If we understand PW method factor as only method bias, then it follows that these two items are purely method effects, but surely they must share some trait variance. In the same vein, items 10 ("losing confidence in yourself") and 11 ("thinking of yourself as a worthless person") load very high in the NW method factor and, as a reviewer pointed out, a likely (post-hoc) explanation is that wording bias are still confounded with a confidence/self-image factor. Therefore, the interpretation of these effects as purely method may be compromised and, accordingly, the interpretation of an overall score for the scale difficult.

The second aim of this study was to examine the relationships between the method factors associated with both NW and PW items and three demographic variables, namely, the sex, age and educational level of the respondents. Regarding the sex, we found a statistically significant, but weak, relationship between PW and sex, so that men were more likely than women to endorse PW items. These results are in line with previous works that, in the context of RSES, have found sex differences in wording effects[56,57]. As for the explanatory role of age on method effects, we found that the relationship between age and the negative wording effect was not statistically significant, which supports previous research using other questionnaires (e.g., self–esteem

scales,[50]; Hospital Anxiety & Depression Scale,[64]). Moreover, our results give support to previous studies which had stated that, in older adults, the strongest method effects would be associated with PW items, rather than NW items[55,58].

As to the educational level, we found that there was not a significant correlation of this variable on the two method factors. This result supports and extends the evidence obtained in Tomás et al.[50], that found that the educational level of the respondents had no effect on the negative method factor using self–esteem questionnaires. This results contradicts previous research on the relationship of the negative wording factor and the educational level/verbal ability with different questionnaires and samples[41,64–69].

Overall, the significant effects of sex and age on trait and method factors point out that women have a worse wellbeing, but this effect is partly modified by a method effect on the positively worded items, whereas the results for age suggest that older respondents have worse well-being and this effect is magnified by a method effect on the positive wording factor. The results on the individual differences related to the demographic variables considered in this study cannot only help to understand the presence of wording method effects but also to identify respondents who are prone to answering PW and NW items differently. In this sense, the relationship that appears as more evident is for the age and sex variables.

Another practical consequence of our study concerns the relationship between the intended measure of the GHQ–12 (i.e., the GPH factor) and other constructs of interest. Several studies have shown that method effects can inflate, deflate or have no effect at all on estimates of the relationship between two constructs (see Podsakoff[70], for a further review of the effects that method biases have on individual measures and on the covariation between different constructs). Thus, it is desirable that both the constructs of interest as well as the effects of method factors, like positive and negative wording, are considered in SEM models as a means of controlling these systematic sources of bias and, thus, avoiding the drawing of inaccurate conclusions about the relationships

between the substantive factors.

Previous research on the GHQ-12 (e.g.[31,36]) has outlined the asymmetry in the participants' responses as a function of the wording of the items, as well as the different responses scales for the positive and negative items. This asymmetry in the participants' responses as a function of the wording of the items is consistent with results from previous research into wording effects for contrastive survey questions[71]. The extent to which the presence of method effects is linked to the asymmetric pattern of responses and/or to the different response scales for the PW and NW items in the GHQ–12 should be examined in future research.

Comparing the current work with previous studies into the factorial structure of the GHQ–12, to our knowledge, this is the first study that, on the one hand, tests a comprehensive set of models including method effects associated with both PW and NW items and also explores some demographic correlates of these method effects. Another strength of this work was the fact that it used a large representative sample of workers, but the results might not be generalized to other specific populations, for example, adolescents and elderly retired people.

Funding

This work was supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

No conflict of interest has been declared by the authors.

Contributorship Statement

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data. MFR and GM: drafted the article. JV and JML: critically revised the draft for important intellectual content. JMT worked in the statistical analysis and interpretation of data. All authors agreed on the final version.

References

1 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: : Oxford University Press 1972.

2 Goldberg DP, Williams P. *A user's guide to the General Health Questionnaire*. Windsor, United Kingdom: : NFER-Nelson 1988.

3 Bhui K, Bhugra D, Goldberg D. Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Soc Psychiatry Psychiatr Epidemiol* 2000;**35**:248–54.

4 Daradkeh TK, Ghubash R, El-Rufaie O. Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychol Rep* 2001;**89**:85–94.

5 Gelaye B, Tadesse MG, Lohsoonthorn V, *et al.* Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J Affect Disord* 2015;**187**:197–202. doi:10.1016/J.JAD.2015.08.045

6 Goldberg DP, Gater R, Sartorius N, *et al.* The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;**27**:191–7.

7 Lundin A, Hallgren M, Theobald H, *et al.* Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health* 2016;**136**:66–74. doi:10.1016/J.PUHE.2016.03.005

8 Rocha K, Pérez K, Rodríguez-Sanz M, *et al.* Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *Int J Clin Heal Psychol* 2011;**11**:125–39.

9 Sanchez-Lopez MP, Dresch V. The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema* 2008;**20**:839–43.

10 Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatr Scand* 1999;**100**:462–8.

11 Tait RJ, French DJ, Hulse GK. Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents. *Aust N Z J Psychiatry* 2003;**37**:374–81.

12 Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol* 1991;**26**:132–8.

13 Campbell A, Knowles S. A confirmatory factor analysis of the GHQ12 using a large Australian sample. In: *European Journal of Psychological Assessment*. 2007. 2–8.

14 Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *Int J Geriatr Psychiatry* 2002;**17**:739–44.

15 French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004;**13**:1–7.

16 Gao F, Luo N, Thumboo J, *et al.* Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health Qual Life Outcomes* 2004;**2**:1–7. doi:10.1186/1477-7525-2-63

17 Mäkikangas A, Feldt T, Kinnunen U, *et al.* The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. *Psychol Assess* 2006;**18**:444–51.

18 Padrón A, Galán I, Durbán M, *et al.* Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Qual Life Res* 2012;**21**:1291–8.

19 Penninkilampi-Kerola V, Miettunen J, Ebeling H. Health and disability: A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scand J Psychol* 2006;**47**:431–40.

20 Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: A

large sample analysis using confirmatory factor analysis. *Psychol Assess* 2005;**17**:231–6.

21 Martin CR, Newell RJ. The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement. *J Psychosom Res* 2005;**59**:193–9. doi:10.1016/j.jpsychores.2005.02.020

22 Tomás JM, Meléndez JC, Oliver A, *et al.* Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores. *Psicológica* 2010;**31**:383–400.

23 Andrich D, Van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychol Med* 1989;**19**:469–85.

24 Gao W, Stark D, Bennett MI, *et al.* Using the 12-item General Health Questionnaire to screen psychological distress from survivorship to end-of-life care: Dimensionality and item quality. *Psychooncology* 2012;**21**:954–61.

25 Glozah FN, Pevalin DJ. Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *J Child Adolesc Ment Heal* 2015;**27**:53–7. doi:10.2989/17280583.2015.1007867

26 Kilic C, Rezaki M, Rezaki B, *et al.* General Health Questionnaire (GHQ-12 & GHQ-28): Psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol* 1997;**32**:327.

27 Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: Reliability, validity and factor structure of the GHQ-12. *Eur Acad Dermatology Venereol* 2001;**15**:410.

28 Werneke U, Goldberg DP, Yalcin I, *et al.* The stability of the factor structure of the General Health Questionnaire. *Psychol Med* 2000;**30**:823.

29 Aguado J, Campbell A, Ascaso C, *et al.* Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish post-partum women. *Assessment* 2012;**19**:517–25.

30 Ye S. Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Pers Individ Dif* 2009;**46**:197–201.

31 Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clin Pract Epidemiol Ment Heal* 2008;**4**:10.

32 Harvey RJ, Billings RS, Nilan KJ. Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *J Appl Psychol* 1985;**70**:461–8.

33 Smith N, Stults DM. Factors defined by negatively keyed items: The results of careless respondents? *Appl Psychol Meas* 1985;**9**:367–73.

34 Abubakar A, Fischer R. The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress Heal J Int Soc Investig Stress* 2012;**28**:248–54.

35 Fernandes HM, Vasconcelos-Raposo J. Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment* 2013;**20**:219–29. doi:http://dx.doi.org/10.1177/1073191112465768

36 Molina JG, Rodrigo MF, Losilla J-M, *et al.* Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychol Assess* 2014;**26**:1031–1037. doi:10.1037/a0036472

37 Motamed N, Edalatian Zakeri S, Rabiee B, *et al.* The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Appl Res Qual Life* 2018;**13**:303–16. doi:10.1007/s11482-017-9522-y

38 Rey JJ, Abad FJ, Barrada JR, *et al.* The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychol Assess* 2014;**26**:1021–30. doi:10.1037/a0036468

39 Romppel M, Braehler E, Roth M, *et al.* What is the General Health Questionnaire-12 assessing? *Compr Psychiatry* 2013;**54**:406–13. doi:10.1016/j.comppsych.2012.10.010

40 Smith AB, Oluboyede Y, West R, *et al.* The factor structure of the GHQ-12: The interaction between item phrasing, variance and levels of distress. *Qual Life Res An Int J Qual Life Asp*

Treat Care Rehabil 2013;**22**:145–52.

41 Marsh HW. Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Dev Psychol* 1986;**22**:37–49.

42 Tomás JM, Oliver A. Rosenberg's Self-Esteem Scale: Two factors or method effects. *Struct Equ Model A Multidiscip J* 1999;**6**:84–98.

43 Gnambs T, Staufenbiel T. The structure of the General Health Questionnaire (GHQ-12): two meta-analytic factor analyses. *Health Psychol Rev* 2018;**12**:179–94. doi:10.1080/17437199.2018.1426484

44 Wang L, Lin W. Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Pers Individ Dif* 2011;**50**:1056–61.

45 Wothke W. Models for multitrait-multimethod matrix analysis. In: Marcoulides GA, Schumacker RE, eds. *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, NJ: : Lawrence Erlbaum Associates 1996.

46 Jöreskog KG. Analyzing psychological data by structural analysis of covariance matrices. In: Atkinson RC, Krantz DH, Luce RD, *et al.*, eds. *Contemporary developments in mathematical psychology. Vol. 2*. San Francisco: : Freeman 1974. 1–56.

47 Markon KE. Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol* 2019;**15**:51–69. doi:10.1146/annurev-clinpsy-050718-095522

48 Reise SP. The Rediscovery of Bifactor Measurement Models. *Multivariate Behav Res* 2012;**47**:667–96. doi:10.1080/00273171.2012.715555

49 Marsh HW, Bailey M. Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models. *Appl Psychol Meas* 1991;**15**:47–70. doi:10.1177/014662169101500106

50 Tomás JM, Oliver A, Galiana L, *et al.* Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Struct Equ*

Model A Multidiscip J 2013;**20**:299–313.

51 Tomás JM, Gutiérrez M, Sancho P. Factorial Validity of the General Health Questionnaire 12 in an Angolan Sample. *Eur J Psychol Assess* 2017;**33**:116–22. doi:10.1027/1015-5759/a000278

52 Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychol Assess* 2010;**22**:366–81.

53 DiStefano C, Motl RW. Self-Esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Struct Equ Model A Multidiscip J* 2009;**16**:134–46.

54 Gana K, Saada Y, Bailly N, *et al.* Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *J Res Pers* 2013;**47**:406–16.

55 Lindwall M, Barkoukis V, Grano C, *et al.* Method effects: The problem with negatively versus positively keyed items. *J Pers Assess* 2012;**94**:196–204.

56 Michaelides MP, Zenger M, Koutsogiorgi C, *et al.* Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Pers Individ Dif* 2016;**97**:13–8. doi:10.1016/J.PAID.2016.03.011

57 Urbán R, Szigeti R, Kökönyei G, *et al.* Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behav Res Methods* 2014;**46**:488–98. doi:10.3758/s13428-013-0391-5

58 Mullen SP, Gothe NP, McAuley E. Evaluation of the factor structure of the Rosenberg Self-Esteem Scale in older adults. *Pers Individ Dif* 2013;**54**:153–7.

59 Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 1996;**70**:810–9. doi:10.1037/0022-3514.70.4.810

60 Catalonian Labor Relations and Quality of Work Department. *Segunda Encuesta Catalana de Condiciones de Trabajo [Second Catalonian Survey of Working Conditions]*. Barcelona: : Author 2012.

61 Lobo A, Muñoz PE. Versiones en lengua española validadas. In: Goldberg D, Williams P, eds. *Cuestionario de Salud General GHQ (General Health Questionnaire)*. Guia para el usuario de las distintas versiones [Guide for the use of the different validated versions in Spanish language]. Barcelona, Spain: : Editorial Masson 1996.

62 Muthén LK, Muthén BO. *Mplus User's Guide*. 6th ed. Los Angeles, CA: : Muthén & Muthén 2011.

63 Browne MW. Asymptotic distribution free methods in the analysis of covariance structures. *Br J Math Stat Psychol* 1984;**37**:62–83.

64 Wouters E, Booysen F le R, Ponnet K, *et al.* Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PLoS One* 2012;**7**:1–7. doi:10.1371/journal.pone.0034881

65 Bors DA, Vigneau FO, Lalande F. Measuring the Need for Cognition: Item polarity, dimensionality, and the relation with ability. *Pers Individ Dif* 2006;**40**:819–28.

66 Chen YH, Rendina-Gobioff G, Dedrick RF. Factorial Invariance of a Chinese Self-Esteem Scale for Third and Sixth Grade Students: Evaluating Method Effects Associated with Positively and Negatively Worded Items. *Int J Educ Psychol Assess* 2010;**6**:21–35.

67 Corwyn RF. The factor structure of global self-esteem among adolescents and adults. *J Res Pers* 2000;**34**:357–79.

68 Rammstedt B, Goldberg LR, Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *J Res Pers* 2010;**44**:53–61.

69 Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *J Pers*

Soc Psychol 2005;**89**:623–42.

70 Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 2012;**63**:539–69.

71 Kamoen N, Holleman B, Mak P, *et al.* Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions. *Discourse Process* 2011;**48**:355–85. doi:10.1080/0163853X.2011.578910

→FILE Figure1.pdf

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

→FILE Figure2.pdf

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

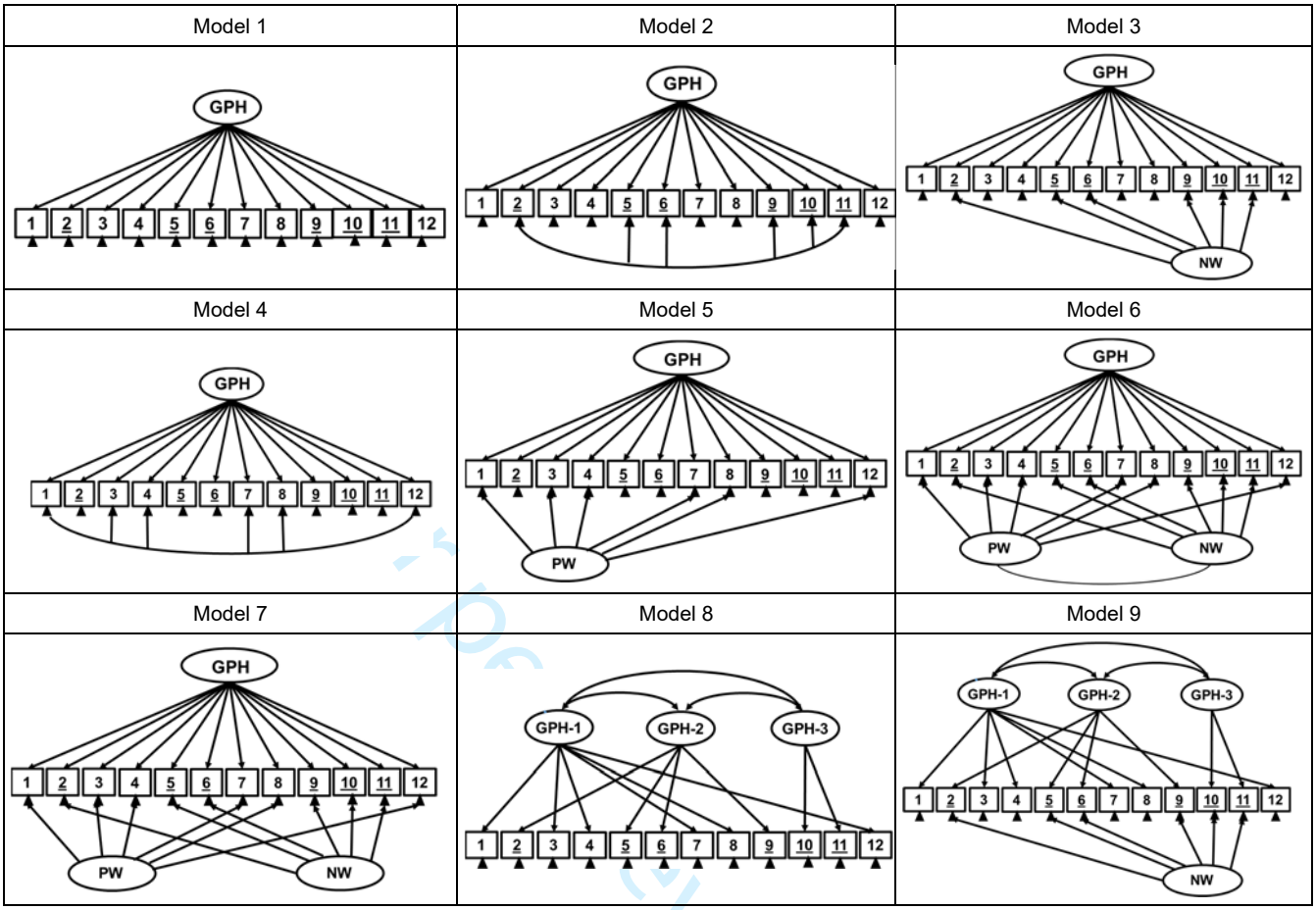

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; GPH–1: Social dysfunction; GPH–2: Anxiety and depression; GPH–3: Loss of confidence; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

Positively worded items

Negatively worded items

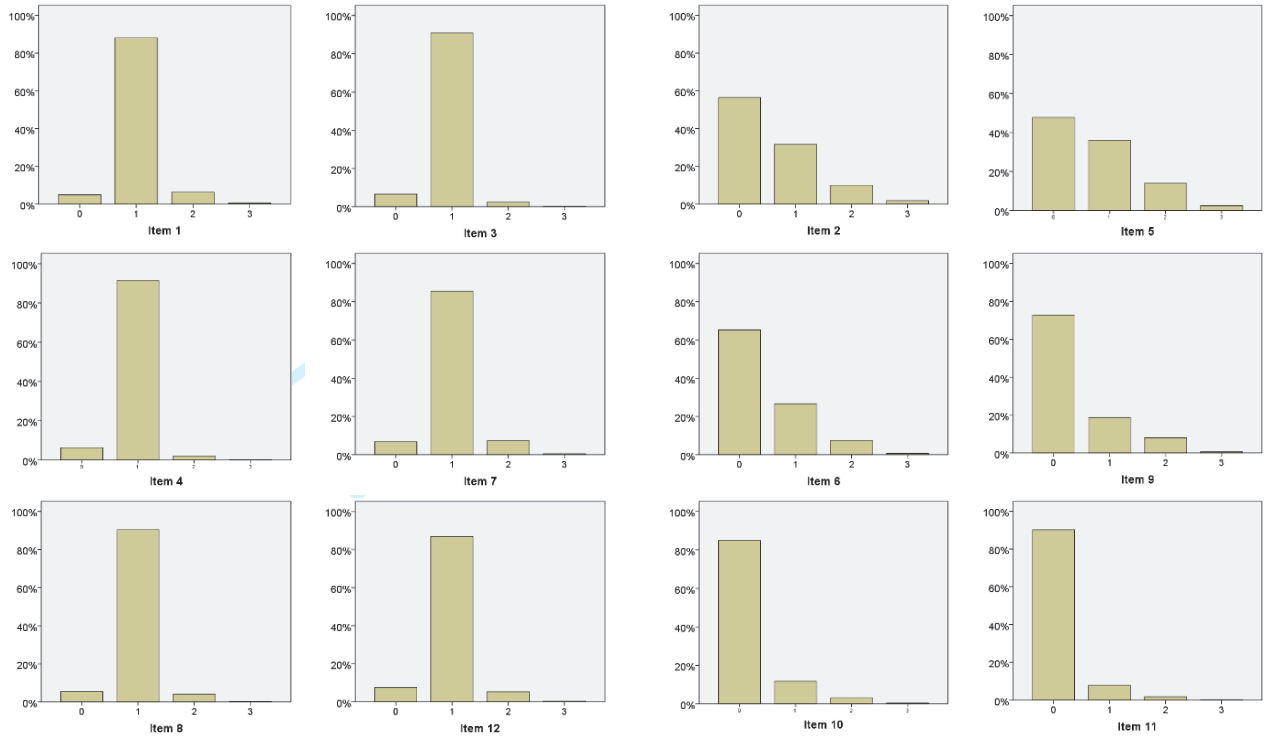

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract P.1 Lines 1-4 (Title)
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found P.3 Lines 1-18 (Abstract)
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported P.5 - P.8 (Introduction)
Objectives	3	State specific objectives, including any prespecified hypotheses P.8 Lines 5-9 (Introduction)
Methods		
Study design	4	Present key elements of study design early in the paper P.8 Lines 13-14 (Method - Participants)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection P.8 Lines 15-16 (Method - Participants)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants P.8 Lines 14-15 (Method - Participants)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable P.9 Lines 2-11 and P.11 Lines 1-5 (Method - Measures)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group P.8 Lines 14-16 and P.9 Lines 2-11 (Method - Measures)
Bias	9	Describe any efforts to address potential sources of bias P.8 Line 14 (Method - Participants)
Study size	10	Explain how the study size was arrived at P.8 Line 13 (Method - Participants)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why P.11 Lines 14-24 and P.12 Line 2 (Method – Statistical Analysis)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding P.11 Lines 7-24 (Method – Statistical Analysis)
		(*b*) Describe any methods used to examine subgroups and interactions P.11 Line 25 and P.12 Lines 1-2 (Method – Statistical Analysis)
		(*c*) Explain how missing data were addressed P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy P.8 Line 13 (Method – Participants)
		(*e*) Describe any sensitivity analyses P.11 Lines 14-24 (Method – Statistical Analysis)
Results		

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed P.8 Lines 16-18 (Method – Participants)
		(b) Give reasons for non-participation at each stage P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
		(c) Consider use of a flow diagram P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders P.8 Table 1 (Method – Participants)
		(b) Indicate number of participants with missing data for each variable of interest P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
Outcome data	15*	Report numbers of outcome events or summary measures P.9 Table 2 (Method - Measures) and P.12 Figure 2 (Method – Statistical Analysis)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included P.9 Table 2 (Method - Measures) and P.12 Table 3 (Results)
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses P.13 Lines 4-14 (Results)
Discussion		
Key results	18	Summarise key results with reference to study objectives P.13 Lines 23-25 and P.14 Lines 1-14 (Discussion)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias P.4 Lines 10-14 (Limitations) and P.15 Lines 20-21 (Discussion)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence P.14 Lines 21-25 and P.15 Lines 1-15 (Discussion)
Generalisability	21	Discuss the generalisability (external validity) of the study results P.15 Lines 16-20 (Discussion)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based P.16 Lines 1-5 (Funding)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative sample of Catalonian workers

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031859.R3
Article Type:	Original research
Date Submitted by the Author:	03-Oct-2019
Complete List of Authors:	Rodrigo, Maria F.; University of Valencia Molina, J. Gabriel; University of Valencia Losilla, Josep-Maria; Autonomous University of Barcelona Vives, Jaume; Autonomous University of Barcelona Tomás, José; University of Valencia
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Research methods, Mental health
Keywords:	psychological health, General Health Questionnaire (GHQ–12), method effects, item wording effects, confirmatory factor analysis

1 **Title**

2 Method Effects Associated with Negatively and Positively Worded Items on the 12–Item General
3 Health Questionnaire (GHQ–12): results from a cross-sectional survey with a representative
4 sample of Catalonian workers
5

6 2. Authors and affiliation

7 Rodrigo, Maria F.
8 Department of Methodology for the Behavioral Sciences
9 University of Valencia
10 Valencia (Spain)
11
12 Molina, J. Gabriel †
13 Department of Methodology for the Behavioral Sciences
14 University of Valencia
15 Valencia (Spain)
16
17 Losilla, Josep-Maria
18 Department of Psychobiology and Methodology of Health Sciences
19 Universitat Autònoma de Barcelona
20 Barcelona (Spain)
21
22 Vives, Jaume
23 Department of Psychobiology and Methodology of Health Sciences
24 Universitat Autònoma de Barcelona
25 Barcelona (Spain)
26
27 Tomás, José M.
28 Department of Methodology for the Behavioral Sciences
29 University of Valencia
30 Valencia (Spain)
31
32 †Deceased September 26, 2014
33

34 3. Word count, excluding title page, abstract, references, figures and tables: 3119
35

36 4. Authors' statements

37 All authors have agreed to authorship in the indicated order.
38 All authors declare that this paper is an original unpublished work and it is not being submitted
39 elsewhere.
40 All authors do not have any financial interests that might be interpreted as influencing the research,
41 and APA ethical standard were followed in the conduct of the study.
42

5. Ethics statement.
The research was not submitted to approval by an institutional review board since this is not a requirement at our universities for this type of study.
Ethics approval was not sought for this study since this was a secondary analysis of anonymized data.

6. Data availability statement
Data is publicly available upon request to the Catalonian Labor Relations and Quality of Work Department.

7. Contact information for the corresponding author:

Jaume Vives
Department of Psychobiology and Methodology of Health Sciences
Universitat Autònoma de Barcelona
Edifici B - Despatx B5b/081
Carrer de la Fortuna / Carrer de Ca n'Altayó. Campus de la UAB
08193 Bellaterra (Cerdanyola del Vallès). Barcelona (Spain)
E-mail: jaume.vives@uab.cat
Phone: +34 93 581 23 31

Abstract

Objective. Recent studies into the factorial structure of the 12–item version of the General Health Questionnaire (GHQ–12) have shown that it was best represented by a single substantive factor when method effects associated with negatively worded (NW) items are considered. The purpose of the present study was to examine the presence of method effects, and their relationships with demographic covariates, associated with positively worded (PW) and/or NW items.

Design. A cross-sectional, observational study to compare a comprehensive set of confirmatory factor models, including method effects associated with PW and/or NW items with GHQ-12 responses.

Setting. Representative sample of all employees living in Catalonia (Spain).

Participants. 3050 participants (44.6% women) who responded the Second Catalonian Survey of Working Conditions.

Results. A confirmatory factor analysis showed that the best–fitting model was a unidimensional model with two additional uncorrelated method factors associated with PW and NW items. Furthermore, structural equation modeling revealed that method effects were differentially related to both the sex and age of the respondents.

Conclusion. Individual differences related to sex and age can help to identify respondents who are prone to answering PW and NW items differently. Consequently, is desirable that both the constructs of interest as well as the effects of method factors are considered in SEM models as a means of avoiding the drawing of inaccurate conclusions about the relationships between the substantive factors.

Keywords

psychological health, General Health Questionnaire (GHQ–12), method effects, item wording

1 effects, confirmatory factor analysis

Strengths and limitations of this study

Strengths

- Sampling quality: A random and large representative sample of workers and face-to-face administration by professional interviewers.
- Comparison of confirmatory models for positively and/or negatively worded items and the use of two different parameterizations.
- There are no previous studies regarding the demographic correlates of wording effects on the GHQ-12.

Limitations

- The different response scale used for the NW items and the PW items in the questionnaire could be a confounding variable.
- The results might not be generalized to other specific populations as, for example, adolescents and elderly retired people.

Introduction

Originally developed by Goldberg[1], the General Health Questionnaire (GHQ) has been widely used as a screening instrument for measuring General Psychological Health (GPH) in both community and non–psychiatric clinical settings[2]. The shortest 12–item version (GHQ–12) is the most popular and has been employed on different settings and in several countries, as well as part of multiple major national health, social wellbeing and occupational surveys, achieving results which underline the fact that it is highly reliable and valid[3–11].

Despite its broad application, the factor structure underlying the responses to the GHQ–12 remains a controversial issue. In this sense, although the GHQ–12 was originally developed as a unidimensional scale, this one–factor latent structure has found little empirical support and some alternative multidimensional models have been proposed as more appropriate. Thus, the one with the most empirical support is the three–factor model proposed by Graetz[5,12–22]. It is important to note that the 6 positively worded (PW) items make up the first factor, whereas the other two factors are made up of the 6 negatively worded (NW) items (see Figure 1, Model 8). On the other hand, the bidimensional model, where the 6 NW and the 6 PW items in the GHQ–12 are grouped into two factors, has also obtained wide support, especially in studies based on exploratory factor analysis [5,10,23–28]. The arguments against these models and in favor of the unidimensional solution are the high correlations between the factors [13] and the low discriminant validity of the factor scores derived from these models[16,29,30].

As Hankins[31] points out, multifactor models may just be the resulting artifact of the inclusion of PW and NW items in the questionnaire and so, the controversy about the factorial structure of the GHQ–12 might relate to the effect of item wording on subjects' response patterns as part of a more general category called 'method[32,33]. Hankins[31] found that, after modeling the wording effects for the NW items, the unidimensional model fitted better than both the two–factor model (NW vs. PW items) and Graetz's three–factor model. Other studies have called into question

the substantive meaning of the GHQ–12 multifactor solutions, suggesting that they might just be an artifact due to the wording effects associated with NW items[29,30,34–40]. See Molina et al.[36] for a deeper review about the dimensionality of GHQ-12.

Some studies about other instruments, however, suggested not only considering the wording effects for the NW items but also for the PW items[41,42]. Regarding GHQ-12, only a recent meta-analysis modelled the presence of method effects for negatively and positively worded items concluding that positively keyed items explained incremental variance beyond a general mental health factor[43].

Therefore, another source of variability in the results about the factor structure of the GHQ-12 could come from the statistical control of method biases, that has been mainly achieved through the correlated traits–correlated methods (CTCM) and the correlated traits–correlated uniquenesses (CTCU) confirmatory factor analysis models. Both procedures have been used in GHQ–12, to deal with method effects applying the CTCM model [i.e., 30,44] the CTCU model [i.e., 29,31,39,40], or both CTCM and CTCU [i.e., 34–37].

To date, we have not found any study about GHQ-12 that analyze the wording effects associated with either PW items alone, nor with NW and PW items simultaneously, comparing both CTCU and CTCM models. There are several multivariate statistical models for analyzing method effect, and among them the CFA based approaches are the most popular ones [45], in particular the CFA with correlated traits and correlated methods (CFA-CTCM) and the CFA with correlated traits and correlated uniqueness (CTCU). On one hand, the CTCM model specifies that indicators' variance can be explained by a linear combination of trait, method, and error effects [46], with trait and method effects specified as latent variables. The CTCM model, when methods are specified independent (uncorrelated), directly translates into the well-known Bifactor model [47,48]. On the other hand, the CTCU model specifies trait factors while method effects are modeled correlating the uniqueness of items (indicators) sharing a common method [49]. Both CTCM and CTCU models

have strengths and shortcomings and therefore are usually employed simultaneously [50]. This work extends the previous work by Molina et al. [36], which compares the fit of the unidimensional model, the multifactor models and the CTCM and CTCU unidimensional models with method effects for only the NW items.

To clarify, Figure 1 (Model 1 to Model 9) shows the 9 CFA models estimated to test the potential method effects associated with either the PW or the NW, or both. Model 1 is a one factor model of general health. This model also works as a baseline model against which to compare other more complex models. Models 2 and 3 are the CTCU and CTCM models that include method effects for the NW items. These were the best fitting models in Molina et al. [36]. Models 4 and 5 are the CTCU and CTCM models including method effects for the PW items. Model 6 is the CTCM model including method factors for both the NW and PW items (a CTCU model with method effects for both PW and NW items was not estimated because it is not identified). Model 7 is a bifactor model with a general trait factor of general health and two method factors associated to NW and PW items. The three factors are independent (uncorrelated). Additionally, and considering the best fitting multidimensional model in Tomás, Gutiérrez and Sancho [51] based on the results by Graetz [12], models 8 and 9 were also tested. Model 8 posited three substantive dimensions: social dysfunction, anxiety and depression, and loss of confidence. Model 9 included an additional method factor associated to NW items. Models considering a method factor associated to PW items made no sense as all PW items were indicators of social dysfunction.

As stressed by Marsh et al. [52], it becomes necessary to consider this comprehensive set of competing models to determine the relative importance and substantive nature of the method effects.

Figure 1

Finally, there has been some research carried out on the demographic correlates of method

effects, such as sex[53–57], age[55,58] or educational level[41,59]. With respect to the GHQ–12, to date, we have not found any studies that analyze demographic correlates of method effects.

Building on the previous studies, the first aim of this study was to overcome the limitation pointed out in Molina et al.[36] and examine method effects associated with both positive and negative wording. The second aim was to further understand the meaning of the method factors; therefore, we evaluated the relationships between the method factors and three covariates (i.e., the sex, age, and educational level) in the framework of a structural equation model (SEM).

Method

Participants

The data used in this study came from the Second Catalonian Survey of Working Conditions[60] and were based on a representative random sample of all employees living in Catalonia (Spain). Data were collected between September and November 2010 by professional interviewers in private households. The sample comprised a total of 3,050 participants who responded to the GHQ–12 included in the survey. Main sociodemographic characteristics of the sample are shown in Table 1.

Table 1. *Main sociodemographic characteristics*.

	M (SD)	n (%)	Range
Gender			
Women		1361 (44.6)	
Age	40.46 (11.19)		17-82
Education			
Incomplete primary studies		90 (3.0)	
Primary studies		541 (17.9)	
Secondary studies. 1st stage		637 (21.0)	
Associate degree		763 (25.2)	
High School		598 (19.8	
Graduate studies		359 (11.9)	
Postgraduate studies		39 (1.3)	

Note. *M* = Mean; *SD* = Standard Deviation

Public Involvement

Respondents were not involved in any stage of the design of the study and were only requested to respond the survey. In the selected households, interviewers identified themselves personally and informed this was an official survey about the working conditions of employed Catalonian people commissioned by the Catalonian Government Work Department.

Results were published on the Catalonian Government Work Department website[60] and are available at https://treball.gencat.cat/ca/ambits/seguretat_i_salut_laboral/publicacions/estadistiques_estudis/ci/ii_ecct/treballadors/

Measures

The GHQ–12 is a self–report scale that contains 6 PW items (e.g. "Have you been able to face up to problems?") and 6 NW items (e.g. "Have you been losing confidence in yourself?"). The GHQ–12 was validated in Spain by Lobo and Muñoz[61]. Table 2 shows the statements of these items in the same order as they were presented in the survey. It must be noted that the GHQ–12 has a different response scale for the PW items (i.e.: *more than usual*; *same as usual*; *less than usual*; and *much less than usual*) and the NW items (i.e.: *not at all*; *no more than usual*; *rather more than usual*; and *much more than usual*). Accordingly, the 4–point scoring scheme was applied in our study so total scores in the GHQ–12 ranged from 0 to a maximum of 36, with higher scores indicating lower levels of GPH.

Table 2. *Descriptive statistics, standardized factor loadings from Model 7 and correlations between the Model 7 factors and the covariates*

			Model 7		
Item	*M*	*SD*	*GPH*	*PW*	*NW*
Item 1. Able to concentrate	1.03	0.37	.42*	.49*	
Item 2. Lost sleep over worry	0.57	0.75	.78*		.07

Item 3. Playing a useful part in things	0.96	0.31	.09*	.59*	
Item 4. Capable of making decisions	0.96	0.30	.14*	.70*	
Item 5. Constantly under strain	0.71	0.79	.83*		.03
Item 6. Could not overcome difficulties	0.44	0.66	.76*		.25*
Item 7. Enjoy day–to–day activities	1.01	0.40	.53*	.55*	
Item 8. Face up to problems	0.99	0.32	.39*	.60*	
Item 9. Feeling unhappy and depressed	0.37	0.66	.78*		.38*
Item 10. Losing confidence in yourself	0.19	0.48	.53*		.70*
Item 11. Thinking of yourself as a worthless person	0.12	0.40	.48*		.72*
Item 12. Feeling reasonably happy	0.99	0.38	.44*	.72*	

Relations between the Model 7 factors and the socio-demographic variables

Sex	.13*	-.08*	-.02
Age	.11*	.08*	.01
Educational level	.00	-.02	-.06

Note. *M* = Mean; *SD* = Standard Deviation; *GPH* = General Health Psychology factor; *PW* = Positive Wording factor; *NW* = Negative Wording factor. *$p < .05$

For the purposes of exploring the correlates of method effects (i.e., item wording effects), we used the following three covariates: (a) sex (0= men and 1= women); (b) age; and (c) educational level, which was measured as a self–reported question with 7 response graduated categories ranging from *incomplete primary studies* to *postgraduate studies*. The educational level was scored as the highest level of education reached.

Statistical Analysis

A set of competing confirmatory factor models were estimated using MPlus 8.3[62]. Figure 1 shows the specification of all these CFA models. The goodness–of–fit indices computed were: the chi–square statistic; the Comparative Fit Index (CFI); the Root Mean Square Error of Approximation (RMSEA) with its 90% confidence interval; and the Standardized Root Mean Square Residual (SRMR). Values greater than 0.95 for CFI and TLI, and lower than 0.06 and 0.08 for RMSEA and SRMR, respectively, are considered to indicate good model fit.

As concerns the estimation of CFA models, most studies into the GHQ–12 factor structure have used maximum likelihood[16,31,35,40,44]. This estimation method relies on several assumptions which should be met to be confident about the results obtained. This is the case of the assumption of multivariate normality which implies, first, that the variables are continuous in nature and, second, that the joint distribution of the variables is normal. The first condition is unlikely to be met with the GHQ–12 Likert–type response data; nor is the second if the variables depart markedly from normality as is the case for the responses to the NW items which were heavily positively skewed (see Figure 2). An alternative when these conditions are not met is to use the weighted least squares (WLS) estimator [63], which has already been used in some studies about the GHQ–12 factor structure[13,18,20,29] and it will be the estimation method used here. Thus, the various CFA models were estimated using Diagonally WLS.

Finally, correlates of the GHQ–12 factors were evaluated using SEM through the inclusion

in the finally selected model of the 3 covariates considered in this study: sex was treated as categorical, whereas age and educational level were treated as continuous variables.

Figure 2

Results

The goodness–of–fit statistics and indices obtained for the 9 models compared here are shown in Table 3.

Table 3. *Fit indexes for the alternative models of the 12–item General Health Questionnaire*

Models	*df*	Chi-	CFI	RMSEA [90% CI]	SRMR
Model 1	54	5378.68	.77	.180 [.176, .184]	.119
Model 2	39	928.099	.96	.086 [.082, .091]	.049
Model 3	48	1345.38	.95	.094 [.090, .059]	.061
Model 4	39	934.690	.96	.087 [.083, .092]	.052
Model 5	48	1275.28	.95	.092 [.087, .096]	.058
Model 6	41	497.520	.98	.060 [.056, .065]	.030
Model 7	42	507.741	.98	.060 [.056, .065]	.030
Model 8	51	1142.88	.95	.084 [.080, .088]	.054
Model 9	45	960.388	.96	.082 [.078, .086]	.049

Notes: Models are specified in Figure 1; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual.

Model 1, with a single factor of general health, and model 8, with three substantive factors, had worse fit than the models that include wording effects. That is, a careful look at fit indexes makes clear that the inclusion of method effects always improves model fit. Indeed, both NW and PW method effects are needed to get the best fitting models. These best fitting models were models 6 and 7. Their fit was practically indistinguishable and, given that they only differ in that model 7 is more parsimonious because constrains method factors correlation to zero, it will be retained as the best representation of the observed data.

An in–depth inspection of the parameter estimates in Model 7 (see Table 2) showed that all factor loadings were statistically significant for the three factors, except for items 2 and 5 in the method factor comprising the NW items.

Finally, a statistical analysis of the relationships between the latent factors in Model 7 and the 3 covariates considered in this study (i.e. sex, age and educational level) was performed through a MIMIC SEM model in which the effects between the 3 latent factors in Model 7 and the 3 covariates were freely estimated, the focus being on the relationships between the method factors and the covariates. The model fit was excellent (RMSEA = .040; RMSEA 90% IC = [.037, .049]; CFI = .99; SRMR = .029). As can be seen in Table 2, the relations of age with the method factors were near to 0 and statistically non–significant for NW items, and positive and significant although small with PW items (.08). Sex was significantly related with the method factor associated with PW items (–.08), whereas the educational level was not significantly related to method factors. Thus, men and women differ in the way they answer PW items, meaning that men are slightly more likely than women to endorse PW items, and method effects associated to PW items also increased by age.

Discussion

This study focused on the examination of the latent structure underlying the responses to the GHQ–12, considering the role of method effects associated with both, positive and negative items wording, and using two alternative parameterizations of the CFA measurement models. What should first be noted is that the studies that have included method effects in the measurement model of the GHQ–12 have been more the exception than the rule in previous research into the factor structure of this questionnaire.

According to the results of the present study, we conclude that the GHQ–12 factor structure is best characterized by introducing latent method factors that capture both the method effects associated with NW and PW items (Model 7). These results support the conclusion from previous

research that the good fit obtained by multidimensional models (mainly the two–factor model and the three–factor Graetz's model) could simply be explained by the artificial grouping of PW and NW items. However, the interpretation of the latent (method) factors as purely integrating method bias due to wording is not straightforward. It is obvious that NW and PW items share the wording. It is also clear that this three bifactor model (one trait and two method factors) fitted the data best. And finally, there is a lot of empirical evidence on these wording effects. However, it is also relevant to discuss the large loadings of many items on the method factors, being these loadings sometimes larger than their loadings in the trait factor. The general factor explains a 52% of the shared variance, but there are some items that deserve careful attention. For example, items 3 ("playing useful part in things") and 4 ("capable of making decisions") had very low loadings on the trait factor. If we understand PW method factor as only method bias, then it follows that these two items are purely method effects, but surely they must share some trait variance. In the same vein, items 10 ("losing confidence in yourself") and 11 ("thinking of yourself as a worthless person") load very high in the NW method factor and, as a reviewer pointed out, a likely (post-hoc) explanation is that wording bias are still confounded with a confidence/self-image factor. Therefore, the interpretation of these effects as purely method may be compromised and, accordingly, the interpretation of an overall score for the scale difficult.

The second aim of this study was to examine the relationships between the method factors associated with both NW and PW items and three demographic variables, namely, the sex, age and educational level of the respondents. Regarding the sex, we found a statistically significant, but weak, relationship between PW and sex, so that men were more likely than women to endorse PW items. These results are in line with previous works that, in the context of RSES, have found sex differences in wording effects[56,57]. As for the explanatory role of age on method effects, we found that the relationship between age and the negative wording effect was not statistically significant, which supports previous research using other questionnaires (e.g., self–esteem

scales,[50]; Hospital Anxiety & Depression Scale,[64]). Moreover, our results give support to previous studies which had stated that, in older adults, the strongest method effects would be associated with PW items, rather than NW items[55,58].

As to the educational level, we found that there was not a significant correlation of this variable on the two method factors. This result supports and extends the evidence obtained in Tomás et al.[50], that found that the educational level of the respondents had no effect on the negative method factor using self–esteem questionnaires. This results contradicts previous research on the relationship of the negative wording factor and the educational level/verbal ability with different questionnaires and samples[41,64–69].

Overall, the significant effects of sex and age on trait and method factors point out that women have a worse wellbeing, but this effect is partly modified by a method effect on the positively worded items, whereas the results for age suggest that older respondents have worse well-being and this effect is magnified by a method effect on the positive wording factor. The results on the individual differences related to the demographic variables considered in this study cannot only help to understand the presence of wording method effects but also to identify respondents who are prone to answering PW and NW items differently. In this sense, the relationship that appears as more evident is for the age and sex variables.

Another practical consequence of our study concerns the relationship between the intended measure of the GHQ–12 (i.e., the GPH factor) and other constructs of interest. Several studies have shown that method effects can inflate, deflate or have no effect at all on estimates of the relationship between two constructs (see Podsakoff[70], for a further review of the effects that method biases have on individual measures and on the covariation between different constructs). Thus, it is desirable that both the constructs of interest as well as the effects of method factors, like positive and negative wording, are considered in SEM models as a means of controlling these systematic sources of bias and, thus, avoiding the drawing of inaccurate conclusions about the relationships

between the substantive factors.

Previous research on the GHQ-12 (e.g.[31,36]) has outlined the asymmetry in the participants' responses as a function of the wording of the items, as well as the different responses scales for the positive and negative items. This asymmetry in the participants' responses as a function of the wording of the items is consistent with results from previous research into wording effects for contrastive survey questions[71]. The extent to which the presence of method effects is linked to the asymmetric pattern of responses and/or to the different response scales for the PW and NW items in the GHQ–12 should be examined in future research.

Comparing the current work with previous studies into the factorial structure of the GHQ–12, to our knowledge, this is the first study that, on the one hand, tests a comprehensive set of models including method effects associated with both PW and NW items and also explores some demographic correlates of these method effects. Another strength of this work was the fact that it used a large representative sample of workers, but the results might not be generalized to other specific populations, for example, adolescents and elderly retired people.

Funding

This work was supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

No conflict of interest has been declared by the authors.

Contributorship Statement

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data. MFR and GM: drafted the article. JV and JML: critically revised the draft for important intellectual content. JMT worked in the statistical analysis and interpretation of data. All authors agreed on the final version.

References

1 Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: : Oxford University Press 1972.

2 Goldberg DP, Williams P. *A user's guide to the General Health Questionnaire*. Windsor, United Kingdom: : NFER-Nelson 1988.

3 Bhui K, Bhugra D, Goldberg D. Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Soc Psychiatry Psychiatr Epidemiol* 2000;**35**:248–54.

4 Daradkeh TK, Ghubash R, El-Rufaie O. Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychol Rep* 2001;**89**:85–94.

5 Gelaye B, Tadesse MG, Lohsoonthorn V, *et al.* Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J Affect Disord* 2015;**187**:197–202. doi:10.1016/J.JAD.2015.08.045

6 Goldberg DP, Gater R, Sartorius N, *et al.* The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;**27**:191–7.

7 Lundin A, Hallgren M, Theobald H, *et al.* Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health* 2016;**136**:66–74. doi:10.1016/J.PUHE.2016.03.005

8 Rocha K, Pérez K, Rodríguez-Sanz M, *et al.* Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *Int J Clin Heal Psychol* 2011;**11**:125–39.

9 Sanchez-Lopez MP, Dresch V. The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema* 2008;**20**:839–43.

10 Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatr Scand* 1999;**100**:462–8.

11 Tait RJ, French DJ, Hulse GK. Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents. *Aust N Z J Psychiatry* 2003;**37**:374–81.

12 Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol* 1991;**26**:132–8.

13 Campbell A, Knowles S. A confirmatory factor analysis of the GHQ12 using a large Australian sample. In: *European Journal of Psychological Assessment*. 2007. 2–8.

14 Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *Int J Geriatr Psychiatry* 2002;**17**:739–44.

15 French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004;**13**:1–7.

16 Gao F, Luo N, Thumboo J, *et al.* Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health Qual Life Outcomes* 2004;**2**:1–7. doi:10.1186/1477-7525-2-63

17 Mäkikangas A, Feldt T, Kinnunen U, *et al.* The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. *Psychol Assess* 2006;**18**:444–51.

18 Padrón A, Galán I, Durbán M, *et al.* Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Qual Life Res* 2012;**21**:1291–8.

19 Penninkilampi-Kerola V, Miettunen J, Ebeling H. Health and disability: A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scand J Psychol* 2006;**47**:431–40.

20 Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: A

large sample analysis using confirmatory factor analysis. *Psychol Assess* 2005;**17**:231–6.

21 Martin CR, Newell RJ. The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement. *J Psychosom Res* 2005;**59**:193–9. doi:10.1016/j.jpsychores.2005.02.020

22 Tomás JM, Meléndez JC, Oliver A, *et al.* Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores. *Psicológica* 2010;**31**:383–400.

23 Andrich D, Van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychol Med* 1989;**19**:469–85.

24 Gao W, Stark D, Bennett MI, *et al.* Using the 12-item General Health Questionnaire to screen psychological distress from survivorship to end-of-life care: Dimensionality and item quality. *Psychooncology* 2012;**21**:954–61.

25 Glozah FN, Pevalin DJ. Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *J Child Adolesc Ment Heal* 2015;**27**:53–7. doi:10.2989/17280583.2015.1007867

26 Kilic C, Rezaki M, Rezaki B, *et al.* General Health Questionnaire (GHQ-12 & GHQ-28): Psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol* 1997;**32**:327.

27 Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: Reliability, validity and factor structure of the GHQ-12. *Eur Acad Dermatology Venereol* 2001;**15**:410.

28 Werneke U, Goldberg DP, Yalcin I, *et al.* The stability of the factor structure of the General Health Questionnaire. *Psychol Med* 2000;**30**:823.

29 Aguado J, Campbell A, Ascaso C, *et al.* Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish post-partum women. *Assessment* 2012;**19**:517–25.

30 Ye S. Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Pers Individ Dif* 2009;**46**:197–201.

31 Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clin Pract Epidemiol Ment Heal* 2008;**4**:10.

32 Harvey RJ, Billings RS, Nilan KJ. Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *J Appl Psychol* 1985;**70**:461–8.

33 Smith N, Stults DM. Factors defined by negatively keyed items: The results of careless respondents? *Appl Psychol Meas* 1985;**9**:367–73.

34 Abubakar A, Fischer R. The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress Heal J Int Soc Investig Stress* 2012;**28**:248–54.

35 Fernandes HM, Vasconcelos-Raposo J. Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment* 2013;**20**:219–29. doi:http://dx.doi.org/10.1177/1073191112465768

36 Molina JG, Rodrigo MF, Losilla J-M, *et al.* Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychol Assess* 2014;**26**:1031–1037. doi:10.1037/a0036472

37 Motamed N, Edalatian Zakeri S, Rabiee B, *et al.* The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Appl Res Qual Life* 2018;**13**:303–16. doi:10.1007/s11482-017-9522-y

38 Rey JJ, Abad FJ, Barrada JR, *et al.* The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychol Assess* 2014;**26**:1021–30. doi:10.1037/a0036468

39 Romppel M, Braehler E, Roth M, *et al.* What is the General Health Questionnaire-12 assessing? *Compr Psychiatry* 2013;**54**:406–13. doi:10.1016/j.comppsych.2012.10.010

40 Smith AB, Oluboyede Y, West R, *et al.* The factor structure of the GHQ-12: The interaction between item phrasing, variance and levels of distress. *Qual Life Res An Int J Qual Life Asp*

Treat Care Rehabil 2013;**22**:145–52.

41 Marsh HW. Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Dev Psychol* 1986;**22**:37–49.

42 Tomás JM, Oliver A. Rosenberg's Self-Esteem Scale: Two factors or method effects. *Struct Equ Model A Multidiscip J* 1999;**6**:84–98.

43 Gnambs T, Staufenbiel T. The structure of the General Health Questionnaire (GHQ-12): two meta-analytic factor analyses. *Health Psychol Rev* 2018;**12**:179–94. doi:10.1080/17437199.2018.1426484

44 Wang L, Lin W. Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Pers Individ Dif* 2011;**50**:1056–61.

45 Wothke W. Models for multitrait-multimethod matrix analysis. In: Marcoulides GA, Schumacker RE, eds. *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, NJ: : Lawrence Erlbaum Associates 1996.

46 Jöreskog KG. Analyzing psychological data by structural analysis of covariance matrices. In: Atkinson RC, Krantz DH, Luce RD, *et al.*, eds. *Contemporary developments in mathematical psychology. Vol. 2*. San Francisco: : Freeman 1974. 1–56.

47 Markon KE. Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol* 2019;**15**:51–69. doi:10.1146/annurev-clinpsy-050718-095522

48 Reise SP. The Rediscovery of Bifactor Measurement Models. *Multivariate Behav Res* 2012;**47**:667–96. doi:10.1080/00273171.2012.715555

49 Marsh HW, Bailey M. Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models. *Appl Psychol Meas* 1991;**15**:47–70. doi:10.1177/014662169101500106

50 Tomás JM, Oliver A, Galiana L, *et al.* Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Struct Equ*

Model A Multidiscip J 2013;**20**:299–313.

51 Tomás JM, Gutiérrez M, Sancho P. Factorial Validity of the General Health Questionnaire 12 in an Angolan Sample. *Eur J Psychol Assess* 2017;**33**:116–22. doi:10.1027/1015-5759/a000278

52 Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychol Assess* 2010;**22**:366–81.

53 DiStefano C, Motl RW. Self-Esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Struct Equ Model A Multidiscip J* 2009;**16**:134–46.

54 Gana K, Saada Y, Bailly N, *et al.* Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *J Res Pers* 2013;**47**:406–16.

55 Lindwall M, Barkoukis V, Grano C, *et al.* Method effects: The problem with negatively versus positively keyed items. *J Pers Assess* 2012;**94**:196–204.

56 Michaelides MP, Zenger M, Koutsogiorgi C, *et al.* Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Pers Individ Dif* 2016;**97**:13–8. doi:10.1016/J.PAID.2016.03.011

57 Urbán R, Szigeti R, Kökönyei G, *et al.* Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behav Res Methods* 2014;**46**:488–98. doi:10.3758/s13428-013-0391-5

58 Mullen SP, Gothe NP, McAuley E. Evaluation of the factor structure of the Rosenberg Self-Esteem Scale in older adults. *Pers Individ Dif* 2013;**54**:153–7.

59 Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 1996;**70**:810–9. doi:10.1037/0022-3514.70.4.810

60 Catalonian Labor Relations and Quality of Work Department. *Segunda Encuesta Catalana de Condiciones de Trabajo [Second Catalonian Survey of Working Conditions]*. Barcelona: : Author 2012.

61 Lobo A, Muñoz PE. Versiones en lengua española validadas. In: Goldberg D, Williams P, eds. *Cuestionario de Salud General GHQ (General Health Questionnaire)*. Guia para el usuario de las distintas versiones [Guide for the use of the different validated versions in Spanish language]. Barcelona, Spain: : Editorial Masson 1996.

62 Muthén LK, Muthén BO. *Mplus User's Guide*. 6th ed. Los Angeles, CA: : Muthén & Muthén 2011.

63 Browne MW. Asymptotic distribution free methods in the analysis of covariance structures. *Br J Math Stat Psychol* 1984;**37**:62–83.

64 Wouters E, Booysen F le R, Ponnet K, *et al.* Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PLoS One* 2012;**7**:1–7. doi:10.1371/journal.pone.0034881

65 Bors DA, Vigneau FO, Lalande F. Measuring the Need for Cognition: Item polarity, dimensionality, and the relation with ability. *Pers Individ Dif* 2006;**40**:819–28.

66 Chen YH, Rendina-Gobioff G, Dedrick RF. Factorial Invariance of a Chinese Self-Esteem Scale for Third and Sixth Grade Students: Evaluating Method Effects Associated with Positively and Negatively Worded Items. *Int J Educ Psychol Assess* 2010;**6**:21–35.

67 Corwyn RF. The factor structure of global self-esteem among adolescents and adults. *J Res Pers* 2000;**34**:357–79.

68 Rammstedt B, Goldberg LR, Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *J Res Pers* 2010;**44**:53–61.

69 Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *J Pers*

Soc Psychol 2005;**89**:623–42.

70 Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 2012;**63**:539–69.

71 Kamoen N, Holleman B, Mak P, *et al.* Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions. *Discourse Process* 2011;**48**:355–85. doi:10.1080/0163853X.2011.578910

→FILE Figure1.pdf

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

→FILE Figure2.pdf

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

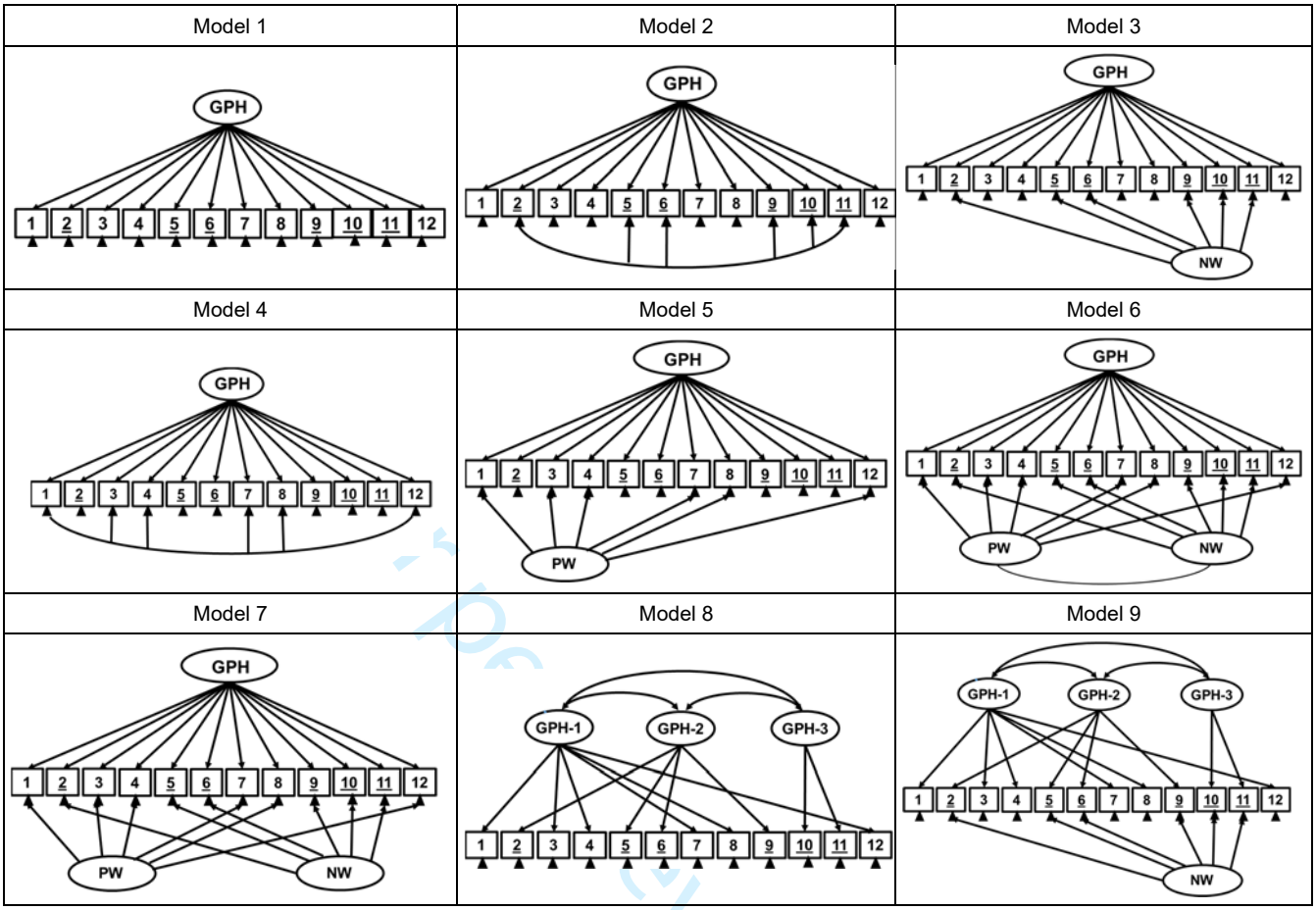

Figure 1. Competing models tested for the 12–Item General Health Questionnaire. Underlined numbers identify negatively worded items. GPH: General Psychological Health factor; GPH–1: Social dysfunction; GPH–2: Anxiety and depression; GPH–3: Loss of confidence; NW: Method factor associated with negatively worded items; PW: Method factor associated with positively worded items.

Positively worded items

Negatively worded items

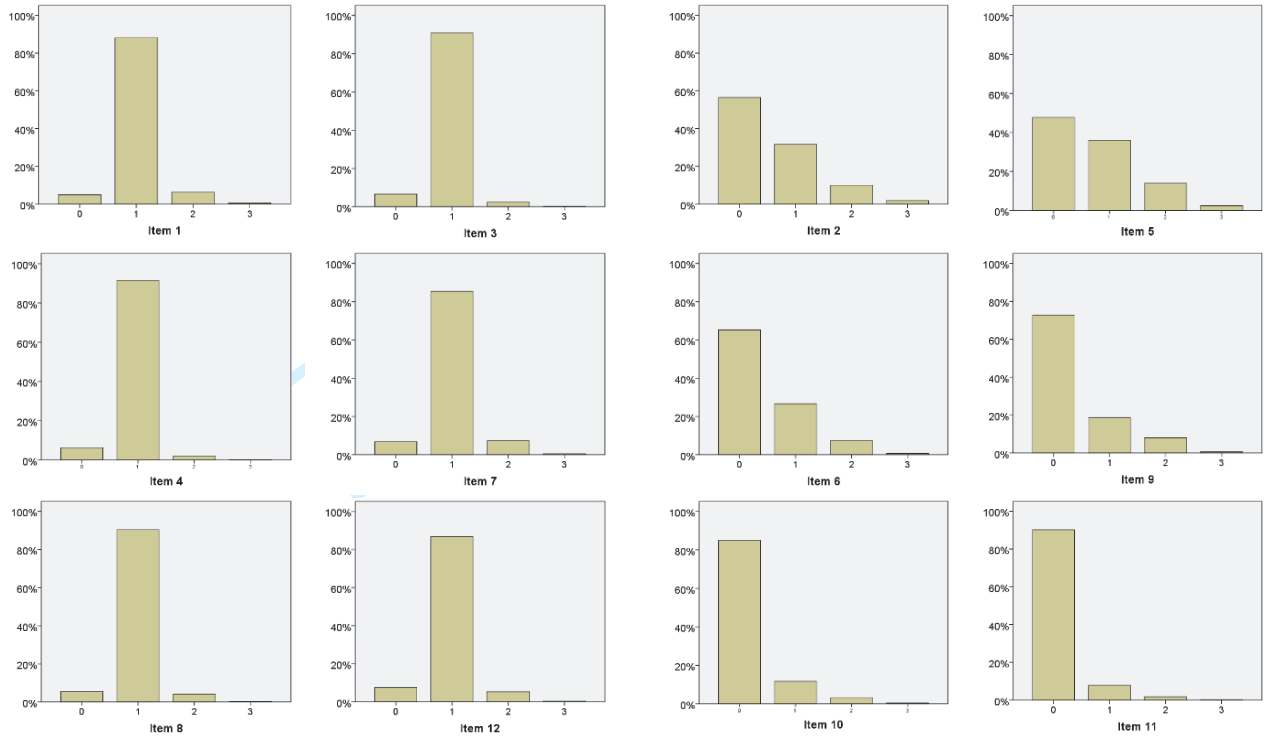

Figure 2. Bar charts of the response distributions for the 12–item General Health Questionnaire. Responses were given on a different 4–point response scale for the positively worded items (0 = *better than usual*, 1 = *same as usual*, 2 = *less than usual*, 3 = *much less than usual*) and for the negatively worded items (0 = *not at all*, 1 = *no more than usual*, 2 = *more than usual*, 3 = *much more than usual*).

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract P.1 Lines 1-4 (Title)
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found P.3 Lines 1-18 (Abstract)
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported P.5 - P.8 (Introduction)
Objectives	3	State specific objectives, including any prespecified hypotheses P.8 Lines 5-9 (Introduction)
Methods		
Study design	4	Present key elements of study design early in the paper P.8 Lines 13-14 (Method - Participants)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection P.8 Lines 15-16 (Method - Participants)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants P.8 Lines 14-15 (Method - Participants)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable P.9 Lines 2-11 and P.11 Lines 1-5 (Method - Measures)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group P.8 Lines 14-16 and P.9 Lines 2-11 (Method - Measures)
Bias	9	Describe any efforts to address potential sources of bias P.8 Line 14 (Method - Participants)
Study size	10	Explain how the study size was arrived at P.8 Line 13 (Method - Participants)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why P.11 Lines 14-24 and P.12 Line 2 (Method – Statistical Analysis)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding P.11 Lines 7-24 (Method – Statistical Analysis)
		(*b*) Describe any methods used to examine subgroups and interactions P.11 Line 25 and P.12 Lines 1-2 (Method – Statistical Analysis)
		(*c*) Explain how missing data were addressed P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy P.8 Line 13 (Method – Participants)
		(*e*) Describe any sensitivity analyses P.11 Lines 14-24 (Method – Statistical Analysis)
Results		

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed P.8 Lines 16-18 (Method – Participants)
		(b) Give reasons for non-participation at each stage P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
		(c) Consider use of a flow diagram P.8 Line 13 (Method – Participants) Previous study cited (reference number 60)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders P.8 Table 1 (Method – Participants)
		(b) Indicate number of participants with missing data for each variable of interest P.8 Lines 13-14 (Method – Participants). No missing data (personal interview)
Outcome data	15*	Report numbers of outcome events or summary measures P.9 Table 2 (Method - Measures) and P.12 Figure 2 (Method – Statistical Analysis)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included P.9 Table 2 (Method - Measures) and P.12 Table 3 (Results)
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses P.13 Lines 4-14 (Results)
Discussion		
Key results	18	Summarise key results with reference to study objectives P.13 Lines 23-25 and P.14 Lines 1-14 (Discussion)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias P.4 Lines 10-14 (Limitations) and P.15 Lines 20-21 (Discussion)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence P.14 Lines 21-25 and P.15 Lines 1-15 (Discussion)
Generalisability	21	Discuss the generalisability (external validity) of the study results P.15 Lines 16-20 (Discussion)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based P.16 Lines 1-5 (Funding)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.