The Structure of the General Health Questionnaire (GHQ-12):

Two Meta-Analytic Factor Analyses

Timo Gnambs

Leibniz Institute for Educational Trajectories

Thomas Staufenbiel

Osnabrück University

Author Note

Timo Gnambs, Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany, Phone: +49 (0)951 863-3420, Email: timo.gnambs@lifbi.de. Thomas Staufenbiel, Institute of Psychology, Osnabrück University, Seminarstrasse 20, 49074 Osnabrück, Germany, Phone: +49 (0)541 969-4512, Email: thomas.staufenbiel@uni-osnabrueck.de.

Correspondence concerning this article should be addressed to Timo Gnambs.

Abstract

The General Health Questionnaire (GHQ-12) is a popular measure of psychological distress. Despite its widespread use, an ongoing controversy pertains to its internal structure. Although the GHQ-12 was originally constructed to capture a unitary construct, empirical studies identified different factor structures. Therefore, this study examined the dimensionality of the GHQ-12 in two independent meta-analyses. The first meta-analysis used summary data published in 38 primary studies (total $N = 76,473$). Meta-analytic exploratory factor analyses identified two factors formed by negatively and positively worded items. The second meta-analysis included individual responses of 410,640 participants from 84 independent samples. Meta-analytic confirmatory factor analyses corroborated the two-dimensional structure of the GHQ-12. However, bifactor modeling showed that most of the variance was explained by a general factor. Therefore, subscale scores reflected rather limited unique variance. Overall, the two meta-analyses demonstrated that the GHQ-12 is essentially unidimensional. It is not recommended to use and interpret subscale scores because they primarily reflect general mental health rather than distinct constructs.

Keywords: mental health, distress, factor analysis, meta-analysis, wording effects

The Structure of the General Health Questionnaire (GHQ-12):

Two Meta-Analytic Factor Analyses

The General Health Questionnaire (GHQ) is a self-report measure of psychological

distress (Goldberg, 1972) that is extensively administered in epidemiological surveys as well

as other community and clinical settings (see Fryers et al., 2004). Particularly, its short form

with 12 items (GHQ-12) exhibits considerable appeal as a quick and unobtrusive screening

instrument to identify people with minor psychological disturbance being at risk of

developing psychiatric disorders (Goldberg & Williams, 1988). Despite its popularity, the

structure of the instrument is still subject to an ongoing debate. Originally, the GHQ-12 was

assumed to capture a single trait. Although some empirical studies supported this assumption

(e.g., Fernandes & Vasconcelos-Raposo, 2012), more frequently some form of

multidimensionality was identified (e.g., Gao et al., 2012; Rey, Abad, Barrada, Garrido, &

Ponsoda, 2014). The latter is sometimes interpreted as a methodological artifact resulting

from wording effects because the GHQ-12 measures positive and negative self-appraisals

with opposing keyed items (e.g., Hankins, 2008a). In contrast, others suggested that the GHQ-

12 measures qualitatively different constructs such as general dysphoria and social

dysfunction (Politi, Piccinelli, & Wilkinson, 1994) and, thus, allows for the interpretation of

different subscale scores. Unfortunately, many of these findings are difficult to evaluate

because they are based on highly selective samples or do not report the fit of competing

models. Therefore, the present study examined the structure of the GHQ-12 within a meta-

analytic structural equation modeling (MASEM) framework (Cheung & Hong, 2017; see also

Gnambs & Staufenbiel, 2016). Two meta-analyses using either summary data or individual

responses from multiple samples evaluated the dimensionality of the scale and scrutinized to

what degree the GHQ-12 variance can be explained by a general factor or more dimensions.

Psychometric Properties of the General Health Questionnaire

The original GHQ consists of 60 items that were subsequently reduced to shorter versions with 30, 28, 20, or 12 items (Goldberg & Williams, 1988). The GHQ-12 includes six positively phrased items (e.g., "Have you been able to concentrate on what you were doing") and six negatively worded items (e.g. "Have you lost much sleep over worry") with four-point response scales (see Table 1). Standard Likert summation yields a global score between 0 and 36, a higher value reflecting more psychological distress (scoring method 0-1-2-3). However, other scoring schemes are also commonly used (see Rey et al., 2014), for example, a dichotomizing that collapses different response categories (0-0-1-1). Studies in different countries have reported a number of good psychometric properties of the GHQ-12 with respect to reliability and validity. Internal consistency reliabilities of the global score ranged from .79 to .91 (Hankins, 2008b; Shevlin & Adamson, 2005), whereas composite reliabilities approached .90 for different scoring methods (Rey et al., 2014). Moreover, test-retest reliabilities fell around .84 after 7 to 14 days (Piccinelli, Bisoffi, Bon, Cunico, & Tansella, 1993), at .79 after 20 days (López-Castedo & Fernández, 2005), and, as could be expected, declined with increasing retest-interval length, $r = .68$ after 12 weeks (Quek, Low, Razack, & Loh, 2001). Validities across 17 studies exhibited a median sensitivity of .84 and a median specificity of .79 (Goldberg et al., 1997). Similar values were found in subsequent studies (Martin & Newell, 2005). Construct validity was also established by means of convergent validity. As expected, the GHQ-12 global score showed a negative correlation with a global quality of life score (Montazeri et al., 2003) and positive associations with depression, state anxiety, and negative affectivity (Tait, French, & Hulse, 2003).

A more controversial issue concerns the factor structure underlying the GHQ-12. Originally, the GHQ-12 was designed as a unidimensional measure. Only a few studies

corroborated this single factor structure (e.g., Fernandes & Vasconcelos-Raposo, 2012). More

support exists for two- and three-dimensional models (e.g., Graetz, 1991; Martin, 1999; Politi

et al., 1994). In an early study, Politi and colleagues (1994) identified two factors that were

labelled 'General Dysphoria' and 'Social Dysfunction'. Whereas the latter included items

relating to enjoying and coping with daily problems, the former reflected general anxiety and

depression. Although the item factor correspondences were not always the same, similar

results were found using exploratory (e.g., Iwata, Okuyama, Kawakami, & Saito, 1988;

Schmitz, Kruse, & Tress, 1999) and confirmatory factor analyses (e.g., Gao et al., 2012;

Gouveia, Barbosa, Andrade, & Carneiro, 2010). However, other two-factor models which are

considerably different from the Politi et al. (1994) model have also been suggested (e.g., Li,

Chung, Chui & Chan, 2009; Vanheule & Bogaerts, 2005). Using latent trait modeling and

data from the GHQ-30, Andrich and van Schoubroeck (1989) demonstrated that positively

and negatively worded items behave differently which can result in a methodological

(artifactual) dual-factor model with all positively worded items loading on one factor and the

negatively framed items on the other. Unfortunately, there is a large overlap between the

substantively meaningful model of Politi and colleagues (1994) and the methodological

artifact model. That is, the items constituting the 'General Dysphoria' factor are positively

worded and the item of the 'Social Dysfunction' largely negatively (item 12 is assigned to

both factors). This makes it difficult to disentangle the substantial psychological construct

model from the methodological artifact model. In the literature, also alternative models with

three factors have been proposed (e.g., Gao et al., 2004; Graetz, 1991; Martin, 1999; Shevlin

& Adamson, 2005), some of which are quite diverse (Campbell, Walker, & Farrell, 2003). For

example, Graetz (1991) found support for a 3-factor structure and distinguished between

'Anxiety' (comprising all positively worded items), 'Anhedonia', and 'Loss of confidence' (a breakdown of the negatively worded items in two factors).

Recently, studies tried to identify the most appropriate structure by comparing fit measures of a wider range of different models using confirmatory factor analyses. Here, too, no consistent structure emerged. For example, Tomás, Gutiérrez, and Sancho (2017) evaluated 20 different structural models of the GHQ-12. Among them were several rarely examined bifactor structures (cf. Reise, 2012) that allowed for facet-specific residual variations beyond a general factor common to all 12 items. These analyses supported Graetz's (1991) 3-factorial model. In contrast, other studies found considerable support for alternative three factor models (e.g., Campbell, Walker, & Farrell, 2003) or even two factor models (e.g., Li, Chung, Chui & Chan, 2009; Rey et al., 2104; Vanheule & Bogaerts, 2005).

Taking into account that the multidimensionality of the GHQ-12 can (at least partially) be explained by artifactual wording effects, newer studies also tried to control for this effect statistically (Hankins, 2008a; Smith, Oluboyede, West, Hewison, & House, 2013; Ye, 2009; Wang & Lin, 2011). The control of this method bias was achieved in confirmatory factor models, which either allowed correlated errors or included additional method factors for differently worded items. In these studies, controlling for wording effects in the GHQ-12 showed a superior fit as compared with models not controlling for different item wording (e.g., Hankins, 2008a; Smith et al., 2013). Studies on the invariance of the factor structure across English and Chinese language versions of the GHQ-12 could also confirm the adequacy of the unidimensional model with wording effects (Chin et al., 2015). However, again the picture is not fully consistent. In some studies, the Graetz (1991) 3-factor model outperformed the single factor model, even when wording effects were controlled for (Abubakar & Fischer. 2012; Tomás et al., 2017). However, sometimes these results are

difficult to compare because models were tested for variants of the GHQ-12 that removed some items (e.g., Wong & O'Driscoll, 2016) or introduced error covariances between items (e.g., Fernandes & Vasconcelos-Raposo, 2012). Additionally, the findings are further complicated by the use of different scoring methods that influence the factor structure and model fit (Rey et al., 2014).

Present Studies

The ongoing controversy surrounding the structure of the GHQ-12 led us to scrutinize its dimensionality from a meta-analytic perspective. Given the prevalent emphasis on replicability in psychological research (e.g., Open Science Collaboration, 2015), we sought to replicate our results in two independent meta-analyses using different data sources and different methodological approaches. Both meta-analyses adopted variants of MASEMs (Cheung & Hong, 2017) to derive pooled correlation matrices between the 12 items included in the GHQ-12 (see also Gnambs & Staufenbiel, 2016). Whereas the first meta-analysis relied on summary data and evaluated the structure of the GHQ-12 using an exploratory approach, the second meta-analysis adopted a confirmatory approach using individual responses from several samples. Moreover, bifactor modeling (Reise, 2012) allowed us to estimate the proportion of common variance explained by a general factor and, thus, to evaluate the meaningfulness of potential subscales. Despite several structural models that have been proposed in the literature (many of which differ on rather few parameters) no consensus as to the adequacy of these models has been reached. Therefore, we adopted meta-analytic exploratory factor analyses to evaluate the GHQ-12 without imposing zero-loading constraints on the loading matrix. This data-driven approach allowed us to evaluate potential (unhypothesized) cross-loadings on several factors. Meta-analytic confirmatory factor analyses were used to capture the multidimensionality of the scale implied by different

structural models described in the literature. Although local misspecifications (e.g., missing factor loadings) can also be identified in confirmatory factor analyses, for example, using modification indices (Saris, Satorra, & van der Veld, 2009), these techniques haven not yet been evaluated within MASEM. Rather, MASEMs are typically judged by model-based goodness-of-fit indices which are known to be sensible to, among others, the item per factor ratio or the average factor loadings (Greiff & Heene, 2017). Therefore, the present study cross-validated the structure of the GHQ-12 in two complementary meta-analyses within an exploratory and a confirmatory framework.

Meta-Analysis I: Exploratory Analyses of Summary Data

Method

Meta-analytic database. Studies reporting on the factor structure of the GHQ-12 were identified in major academic (PsycINFO, Psyndex, EconLit, Business Source Complete, ERIC, SocINDEX, Medline, Scopus, Web of Science, ProQuest Dissertations & Theses Database) and non-academic databases (Google Scholar, Researchgate.net). Using the Boolean expression *general health questionnaire* AND (*exploratory factor analysis* OR *principal components* OR *correlation matrix*) these searches identified in October 2017 a total of 4,668 potential studies. After reviewing the title and the abstracts of these studies, 163 studies were further examined for inclusion in the meta-analytic database. Studies were retained according to the following criteria: (a) The study administered the 12 items included in the GHQ-12. We also considered longer versions of the GHQ as long as they subsumed all items of the GHQ-12. (b) The items were accompanied by their original four-point response scale and (c) used Likert coding of the responses (0-1-2-3). Because factor analyses of Pearson correlations among dichotomous variables typically result in distorted factor solutions (e.g., Kubinger, 2003), we did not include studies that adopted the dichotomous

scoring method (0-0-1-1). (d) The study reported the results of an exploratory factor analysis or provided a full correlation matrix between the 12 items. (e) In case of oblique factor rotations, we only considered studies that also reported the respective factor correlations. (f) Moreover, we excluded factor pattern matrices with an excessive number of missing values (i.e., more than 50%). (g) Finally, one study (Gouveia et al., 2010) was excluded because it reported a nonpositive definite correlation matrix. The results of this search and screening process including a list of excluded studies are summarized in the supplemental material. In total, we identified 38 studies reporting on 45 independent samples that met our inclusion criteria.

   Coding process. The authors developed a coding protocol (see supplemental material) for the extraction of relevant information from each publication that defined all variables and provided guidelines regarding the range of potential values. Two focal statistics were extracted from each study: If a study reported the correlations between the 12 items of the GHQ-12, we noted the respective correlation matrix. Otherwise, we retrieved the factor loadings and the respective factor correlations. In cases where different factor solutions were available for a given sample, we used the factor loading pattern including the largest number of factors. In addition, we extracted descriptive information on the sample (e.g., sample size, country, mean age, percentage of female participants), the publication (e.g., publication year), and the reported factor analysis (e.g., factor analytic method, type of rotation). All codings were conducted by the first author. To evaluate the coding process, 12 randomly selected studies (including about 30% of all samples) were independently coded a second time by a graduate student in psychology. For continuous variables (e.g., factor loadings) intercoder agreement was quantified using two-way intraclass coefficients (ICC; Shrout & Fleiss, 1979); for categorical variables (e.g., factor analytic method) we computed Cohen's (1960) Kappa κ.

According to prevalent guidelines (see LeBreton & Senter, 2008) intercoder agreement can be considered strong for values exceeding .70 and excellent for values greater than .90. The intercoder reliability was ICC = .97, 95% CI [.966, .977] for the factor loadings and ICC = 1.00, 95% CI [1.00, 1.00] for the factor correlations. The remaining variables (e.g., sample size, factor analytic method) had ICCs or Cohen's κ of 1.00. The first author resolved disagreements by revisiting the respective study.

Meta-analytic procedure. The Pearson product-moment correlations between the 12 items of the GHQ-12 were used as effect size measures. Eleven samples reported the respective correlation matrix, whereas 34 samples reported only factor pattern matrices from exploratory factor analyses. For the latter, we calculated the implied correlations between the GHQ-12 items (see indirect method in Gnambs & Staufenbiel, 2016). In eight cases, only partial factor pattern matrices were available because small loadings (e.g., values falling below .40) were not reported. For these matrices, a value of 0 was imputed for the missing factor loadings. Monte Carlo simulations indicated that this approach results in unbiased estimates of the salient factor loadings (Gnambs & Staufenbiel, 2016).

The factor structure of the GHQ-12 was examined with a variant of two-step MASEM (see Cheung & Hong, 2017). In the first step, the item-level correlation matrices were pooled using a multivariate random-effects meta-analysis. Following Cheung (2013), we adopted a structural equation modeling (SEM) framework with a maximum likelihood estimator. In the second step, the thus derived pooled correlation matrix was submitted to an exploratory weighted least square factor analysis. As suggested by Cheung and Chan (2005), the asymptotic sampling covariance matrix of the pooled correlations was used as weight matrix for these analyses. In addition to a direct oblimin rotation ($\delta = 0$; Bernaards & Jennrich, 2005), we also performed a target rotation to a partially specified bifactor structure (Browne,

1972) to disentangle scale-specific factors from a potential general factor underlying all items of the GHQ-12. A diverse set of criteria were used to decide on the number of factors to retain. These included eigenvalue-based criteria such as Kaiser's (1960) rule and Horn's (1965) parallel analysis, Velicer's (1976) minimum average partial test, as well as model fit indices such as the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992). The robustness of the identified factor structure was evaluated in sensitivity analyses that repeated the meta-analysis within various subgroups. The similarity of the factor structures across these subgroups was quantified using coefficients of congruence for individual factors (Tucker, 1951) and coefficients of congruence for complete factor loading matrices (Gebhardt, 1968). Values between .85 and .94 indicate fair similarity, whereas factor structures with values of .95 or above can be considered equal (Lorenzo-Seva & ten Berge, 2006).

Statistical software and data availability. The correlations were pooled using the *metaSEM* software version 0.9.16 (Cheung, 2015) in *R* version 3.4.2. The factor analyses were conducted using routines based on the *psych* package version 1.7.8 (Revelle, 2017) and the *GPArotation* package version 2014-11-1 (Bernaards & Jennrich, 2005). To promote transparency and reproducibility of our analyses (see Nosek et al., 2015), all coded data and analyses scripts are provided in an online repository at http://osf.io/z5c4q/.

Results

Study characteristics. The meta-analysis included 45 independent samples that were published between 1983 and 2016 (*Mdn* = 2006). Each sample comprised of about *Mdn* = 446 participants (total *N* = 76,473; *Min* = 125; *Max* = 8,978) with approximately 54% women and a reported mean age of 36.87 years (*SD* = 16.04). The studies were conducted in 28 different countries around the world, with most samples coming from England (13%), Brazil (11%),

and Japan (9%). The samples primarily administered the GHQ-12 (80%), whereas the rest

received longer versions including either 20 items (4%) or 30 items (16%). The factor

analyses of the GHQ-12 predominantly extracted two factors (84%); the remaining samples

reported three factor solutions. The characteristics of each individual sample are also

summarized in the supplemental material.

 Meta-analytic factor analyses. The homogeneity of the correlation matrices was

examined using a fixed-effects model. The respective fit indices (CFI = .80, RMSEA = .12,

SRMR = .12) did not support the assumption of homogenous correlation matrices across

samples. Therefore, we selected a random-effects model. The pooled correlations for the 12

items of the GHQ-12 (see supplemental material) ranged between .18 and .52 (Mdn = .30),

whereas the respective random variances fell at Mdn = .009 (Min = .004, Max = .023). A

diverse set of decision criteria suggested the extraction of two factors: (a) The first two

unrotated eigenvalues exceeded 1 (λ_1 = 4.52 and λ_2 = 1.37), whereas the third did not (λ_3 =

0.81). (b) Velicer's (1976) minimum average partial criterion for one to four factor solutions

fell at {.020, .020, .032, .049} and thus reached a minimum at one or two factors. (c) The

RMSEA indicated a good model fit (i.e., a RMSEA < .05; Browne & Cudeck, 1992) for two

factors, $RMSEA_2$ = .04, but not for a single factor, $RMSEA_1$ = .09. In contrast, Horn's (1965)

parallel analysis suggested the extraction of three factors. Because most of these criteria

pointed at two substantial factors, we conducted an exploratory factor analysis with oblique

rotation extracting two factors. The respective results are summarized in Table 1. The two

factors closely mirrored the multidimensional model introduced by Andrich and van

Schoubroeck (1989) that separates the positively and negatively keyed items into distinct

facets. On each factor six items had salient loadings, $M(|\lambda|)$ = .60, whereas the other items

exhibited minor cross-loadings, $M(|\lambda|) = .08$. The two factors were substantially correlated at $r = .61$.

Bifactor modeling. Given the correlated factor structure, we examined to what degree the item variances could be explained by a general factor underlying all items of the GHQ-12. To this end, we conducted another exploratory factor analysis with an orthogonal target rotation toward a partially specified bifactor structure (Browne, 1972). The bifactor structure included a general factor for all items and two specific factors for the differently keyed items. The three latent factors were uncorrelated. The general factor can be interpreted as general distress, whereas the specific factors capture the residual variance due to the positively or negatively worded items. The respective results are summarized in Table 1. All items had loadings greater than .40 on the general factor, $M(|\lambda|) = .56$. In contrast, no item had salient loadings ($\lambda > .40$) on the specific factors, $M(|\lambda|) = .18$. Moreover, more than half of the common variance in each item was explained by the general factor (see last column in Table 1). Similarly, about 79% of the explained common variance was attributable to the general factor, whereas the specific factor for the positively and negatively worded items captured 16% and 5%, respectively. Thus, for a large part, the responses to the GHQ-12 were dominated by a single general factor.

Sensitivity analyses. The robustness of the identified factor structure was studied by repeating the meta-analytic bifactor analysis within various subgroups of samples and examining the similarity of the resulting factor structures. We selected three criteria and compared meta-analytic factor structures[1] derived from (a) reported correlation matrices ($k = 11$, $N = 21,715$), full factor loading matrices ($k = 26$, $N = 43,068$), and loading matrices with imputed missing values ($k = 8$, $N = 11,690$), (b) the GHQ-12 ($k = 36$, $N = 61,932$) and longer GHQ versions including either 20 or 30 items ($k = 9$, $N = 14,541$), and (c) English ($k = 10$, N

= 33,947), Spanish ($k = 7$, $N = 4,972$), Portuguese ($k = 5$, $N = 8,713$), and Japanese ($k = 4$, $N = 6,036$) language versions. The overall factor structures exhibited high similarity across these criteria (see supplemental material); the factor structure congruence coefficients fell between .96 and .99 ($Mdn = .98$). Particularly, the general factor was robustly replicated across the examined subgroups, $Mdn = 1.00$ ($Min = .99$, $Max = 1.00$); in contrast, the specific factors showed somewhat larger variability ($Min = .79$, $Max = .99$).

Meta-Analysis II: Confirmatory Analyses of Individual-Participant Data

The second meta-analysis extends the previous study on four central accounts: First, instead of summary statistics the present study focuses on individual responses of participants (see Debray et al., 2015). Thus, no potentially biasing reconstructions from incomplete factor loading matrices are necessary. Second, the study relied on participants from a single cultural and language group to avoid potential distortions resulting from imperfect test adaptations. Third, we used only representative samples from large-scale assessments to minimize sampling error and identify a common factor pattern for a given population. Fourth, the previously identified factor structure of the GHQ-12 was tested using a confirmatory approach. Thus, the study intends to replicate the previous results in an individual-participant meta-analysis using a new data source and adopting a different analytical approach.

Method

Meta-analytic database. Individual participant data for the GHQ-12 were retrieved from the *UK Data Archive* (http://www.ukdataservice.ac.uk), a non-profit data catalogue for social, health, and economic surveys conducted in the United Kingdom, using the search term *general health questionnaire*. A sample was included in the meta-analysis if it (a) administered the 12 items of the GHQ-12, (b) in its English language version, (c) accompanied by their original four-point response scales, and (d) drew a representative

sample from the population of the United Kingdom or one of its countries. This search process identified 84 independent samples from several large-scale health and social surveys in England, Scotland, and Northern Ireland. A full list of all included samples is given in the supplemental material.

Meta-analytic procedure. The structure of the GHQ-12 was evaluated by two-step MASEM (Cheung & Hong, 2017). In the first step, we calculated the correlation matrix for the 12 items within each sample (see Cheung & Jak, 2016). Negatively worded items were reverse coded. The correlation matrices for each sample are available at http://osf.io/z5c4q/. As in the previous meta-analysis, these correlation matrices were pooled across samples using SEM with maximum likelihood estimation. In the second step, several confirmatory factor models were fitted to the pooled correlation matrix using a weighted least square estimator. Again, the asymptotic sampling covariance matrix of the pooled correlations was used as weight matrix for these analyses (Cheung & Chan, 2005). Both analyses steps were conducted with the *metaSEM* software version 0.9.16 (Cheung, 2015). The fit of these models was evaluated in line with conventional criteria (cf. Schermelleh-Engel, Moosbrugger, & Müller, 2003) using the *Comparative Fit Index* (CFI), the *Standardized Root Mean Square Residual* (SRMR), and the RMSEA. Models with a CFI $\geq$ .95, a RMSEA $\leq$ .08, and a SRMR $\leq$ .10 were interpreted as "acceptable" and models with CFI $\geq$ .97, RMSEA $\leq$ .05, and SRMR $\leq$ .05 as "good" fitting.

Examined factor models. Different structural models were evaluated that have been frequently used in previous research. All models included unconstrained factor loadings and uncorrelated item uniquenesses. The latent factor variances were fixed to 1 for model identification.

In line with the original construction rationale of the GHQ-12 (Goldberg, 1972), *Model 1* included a single factor explaining the covariances between all items. In contrast, *Model 2* additionally acknowledged potential wordings effects (see Ye, 2009; Wang & Lin, 2011). Thus, we estimated a general factor for all items and an orthogonal specific factor for the negatively worded items (see Figure 1). Sometimes, these types of models are also termed nested factor models (Schulze, 2005) or bifactor-(*S*-1) models (Eid, Geiser, Koch, & Heene, 2017). *Model 3* followed Andrich and van Schoubroeck (1989) and specified two correlated latent factors for the positively (1, 3, 4, 7, 8, 12) and negatively worded items (2, 5, 6, 9, 10, 11; see Figure 1). These factors have either been interpreted as representing wording effects (Hankins, 2008ab) or qualitatively different types of mental health, general dysphoria and social dysfunction (Politi et al., 1994). Because these factors were typically correlated, we also estimated a bifactor structure (see Reise, 2012) to disentangle the effects of a general factor from specific factor influences. Thus, we modeled a general factor common to all 12 items and two orthogonal specific factors for the differently worded items (Model 3b in Figure 1). *Model 4* was introduced by Graetz (1991) and included three correlated factors representing anxiety (2, 5, 6, 9), social dysfunction (1, 3, 4, 7, 8, 12), and loss of confidence (10, 11). An alternative three-factor solution was suggested by Martin (1999). Thus, *Model 5* specified three correlated factors reflecting depression (6, 9, 10, 11, 12), stress (2, 5, 7), and successful coping (1, 3, 4, 8). Again, these models were also estimated with a bifactor structure to separate general and specific factor effects.

Results

The meta-analysis included 84 independent samples that were surveyed between 1987 and 2013 (*Mdn* = 2003). These samples included *N* = 410,640 participants (57% women) in the age from 13 to 100 years (*M* = 45.05; *SD* = 19.55). The ICCs for all items of the GHQ-12

were very small (all ICCs ≤ .014). Thus, most of the variance in the observed item scores was

a result of individual differences between participants and not between samples. Accordingly,

meta-analytic models including random effects for the correlations between the 12 items did

not converge. In contrast, a fixed-effects model indicated a satisfactory fit (CFI = .99,

RMSEA = .03, SRMR = .04) and, thus, homogenous correlation matrices across samples. The

pooled correlation matrix of the GHQ-12 is given in the supplemental material. The

correlations ranged from .22 to .68 (*Mdn* = .38).

Meta-analytic factor analyses. The structure of the GHQ-12 was examined by fitting

different confirmatory models to the pooled correlation matrix. The respective fit statistics in

Table 2 corroborated the findings of our first meta-analysis. The unidimensional model

clearly exhibited an unsatisfactory fit (CFI = .89, SRMR = .15, RMSEA = .11). In contrast,

most multidimensional models showed at least acceptable fits. However, the oblique three-

factor model suggested by Martin (1999) demonstrated an inferior fit as compared to the other

models; the respective bifactor formulation even failed to converge. The best fit in terms of

the information criteria represented the bifactor formulation of Andrich and van Schoubroeck

(1989) that acknowledged different wording effects (CFI = .97, SRMR = .04, RMSEA = .05).

The standardized factor loadings are given in Figure 1. All items had loadings greater than .40

on the general factor, $M(|\lambda|) = .63$ (*Min* = .42, *Max* = .84). But, also the specific factor for the

positively worded items exhibited substantial loadings, $M(|\lambda|) = .44$ (*Min* = .32, *Max* = .55). In

contrast, the specific factor for the negatively worded items had a rather unclear loading

pattern, $M(|\lambda|) = .22$ (*Min* = .07, *Max* = .39). Together, the two specific factors explained

about 23% of the common variance, whereas most of the explained common variance (77%)

was attributable to the general factor (see Table 3). The total score reliability, that is, the

proportion of variance in GHQ-12 scores accounted for by the general factor, was $\omega_H = .85$.

To evaluate the meaningfulness of subscale scores in the GHQ-12, we also calculated omega hierarchical subscale ($\omega_{H.S}$; see Rodriguez, Reise, & Haviland, 2016) for the positively and negatively keyed items. This reflects the proportion of unique variance in the subscale scores (reliability) after accounting for the general factor. $\omega_{H.S}$ was estimated as .39 for the positively worded items and .01 for the negatively worded items indicating that both subscales reflected negligible unique variance and primarily represented the general factor. Thus, in line with the previous meta-analyses, the responses to the GHQ-12 seemed to be dominated by a single general factor.

Replicability of meta-analytic factor structure. The robustness of the identified factor solution was evaluated by comparing the pooled correlation matrices from the two meta-analyses. On average, the pooled correlations derived in the second meta-analysis were all larger than the respective correlations from the first meta-analysis, $M(\Delta r) = .09$ ($SD = .04$). However, as summarized in Table 2, the confirmatory factor analyses fitted to the pooled correlation matrix from the first meta-analysis replicated the previously reported results: Multidimensional models outperformed the single factor model; albeit, again the Martin (1999) model provided an inferior fit. Moreover, bifactor specifications fitted better than comparable correlated trait models. Again, the GHQ-12 was dominated by a general factor explaining between 69% and 75% of the common variance (see Table 3), whereas specific factors were less clearly represented (2% to 30%). Subgroup analyses for different language versions replicated these results (see supplemental material). Finally, a multi-group analysis for the bifactor model of Andrich and van Schoubroeck (1989) showed configural measurement invariance across the two meta-analyses ($\chi^2 = 43,650$, $df = 84$, CFI = .0971, RMSEA = .033). Placing equality constraints on the factor loadings did not result in a drop of the CFI below what is usually considered acceptable ($\Delta\chi^2 = 1,735$, $\Delta df = 12$, ΔCFI < .002;

Meade, Johnson, & Braddy, 2008; Khojasteh & Lo, 2015) and, thus, confirmed metric

measurement invariance.

Discussion

For decades, the GHQ-12 has dominated mental health screenings in applied research

and clinical practice (Fryers et al., 2004). It is surprising that after more than 40 years a

fundamental debate on the structure of the GHQ-12 has not been resolved. Empirical studies

frequently identified unidimensional as well as various multidimensional factor solutions

(e.g., Fernandes & Vasconcelos-Raposo, 2012; Gao et al., 2012; Rey et al., 2014). However,

different sample characteristics, language versions, and analyses methods adopted in these

studies made it difficult to find a consensus. To reconcile these conflicting results, we

presented two meta-analyses that systematically examined the factor structure of the GHQ-12

across samples. These analyses provided three central findings. First, the GHQ-12 is not

strictly unidimensional but also reflects wording effects (see also Hankins, 2008a).

Particularly, positively keyed items explained incremental variance beyond a general mental

health factor. Thus, latent variable modeling needs to acknowledge these dependencies to

properly account for the covariance structure of the GHQ-12. Second, the bifactor structure

with wording effects was rather robust and replicated across different language versions.

Particularly, the general factor and the specific factor pertaining to positively worded items

were highly similar across English, Spanish, Portuguese, and Japanese translations of the

instrument. In contrast, the specific factor for negatively worded items showed more

variability across language versions. Thus, negatively worded items seem to reflect some

form of language-specific variance such as cross-cultural differences in response styles

(Johnson, Kulesa, Cho, & Shavitt, 2005). Finally, bifactor modeling revealed that a single

dominant factor accounted for most of the item variance. In contrast, subscale-specific

variance associated with the wording of the items was rather negligible. Overall, the GHQ-12 seems to represent an essentially unidimensional instrument with spurious secondary dimensions reflecting the wording of the items.

Implications for Applied Measurement

It is not uncommon for many psychological measures used in applied practice to capture a dominant general factor, while also reflecting some minor secondary dimensions (Reise, Moore, & Haviland, 2010) that capture, for example, facet-specific variance (e.g., Henry & Crawford, 2005; Vasconcelos-Raposo, Fernandes, & Teixeira, 2013) or systematic response styles (e.g., Marsh, 1996). Similar, the GHQ-12 is not strictly unidimensional, but also reflects systematic residual variance beyond a general distress (or, reverse coded, mental health) factor. Because these residual variances pertained to differently worded items, this pattern can be interpreted as an expression of specific response styles such as acquiescence (Hankins, 2008a). Thus, the multidimensionality of the GHQ-12 seems to reflect method-specific variance that needs to be controlled for in latent variable analyses modeling responses to the 12 items. In practice, pronounced multidimensionality is problematic if composite scores (e.g., sum scores across all items) are used because these reflect a blend of different latent traits. However, for the GHQ-12 these secondary dimensions seem to be less influential; more than 75% of the explained variance was attributable to the general factor. Thus, applied researchers are likely to introduce a negligible bias in their analyses if they adopt composite scores and ignore wording effects. On the other hand, these results also cast doubts on the usefulness of subscale scores calculated separately for negatively and positively worded items (or other facet models; e.g., Graetz, 1991). In our analyses, respective subscales were highly correlated ($r = .80$), and, thus shared a large proportion of variance. This was also reflected in rather low reliability estimates showing rather limited unique variance captured

by subscales for negatively and positively worded items. Because subscales in the GHQ-12 primarily reflect general factor variance and to a lesser degree unique variance, it does not seem advisable to use these scores in substantial analyses. Indeed, comparative analyses showed that subscale scores rarely exhibited substantially different associations with criterion variables as compared to composite scores for the entire scale (e.g., Aguado et al., 2012; Gao et al., 2004; Shevlin & Adamson, 2005). Researchers interested in a more fine-grained differentiation of mental health would likely be better served with longer versions of the GHQ that exhibit clearer facet structures (see, for example, Klainin-Yobas & He, 2014, on the GHQ-30) or alternative instruments such as the Short-Form 36 (SF-36) Health Survey (Anagnostopoulos, Niakas, & Tountas, 2009).

Cautionary Notes and Outlook

The present studies relied on summary statistics pooled across multiple independent samples to scrutinize the structure of the GHQ-12. Accordingly, these results refer to the factor structure in an average sample (in terms of, for example, sociodemographic, cultural, or psychological characteristics of the included respondents). It is conceivable that specific sample characteristics such as individual differences in reading competences (Gnambs & Schroeders, 2017) or random responding (Huang, Liu, & Bowling, 2015) might contribute to ambiguous factor structures in a given sample that deviates from the presented results. Therefore, future research is encouraged to identify moderating influences that might contribute to the multidimensionality in psychological measures. For example, our results were limited to the Likert scoring method of the GHQ-12 and do not necessarily extend to different scoring schemes (see Rey et al., 2014). Similar, group comparisons require a coherent measurement of mental health across, for example, different assessment contexts (e.g., paper versus computerized tests; cf. Gnambs & Kaspar, 2017), measurement occasions

(Mäkikangas et al., 2006), cultural settings (Romppel et al, 2017), or respondent groups (e.g.,

clinical versus community samples). Moreover, we also want to emphasize that the

comparable factor structure identified in different language versions of the GHQ-12, does not

relieve researchers from demonstrating measurement invariance in the specific sample at

hand. Finally, we hope to see more research that demonstrates the incremental validity of

potential subscale scores beyond a general factor before using and interpreting these scales.

Conclusion

Although the GHQ-12 is not strictly unidimensional, specific factors associated with

the item wording explain limited and contrasting unique variance beyond a general factor.

Therefore, composite scores are likely to exhibit only a minor bias resulting from ignored

multidimensionality. In contrast, it is not recommended to use and interpret subscale scores

because they primarily reflect general mental health rather than distinct constructs.

References

Abubakar, A., & Fischer, R. (2012). The factor structure of the 12-item General Health

   Questionnaire in a literate Kenyan population. *Stress and Health, 28*, 248-254.

   doi:10.1002/smi.1420

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., & Luciano, J. V.

   (2012). Examining the factor structure and discriminant validity of the 12-item

   General Health Questionnaire (GHQ-12) among Spanish postpartum women.

   Assessment, 19, 517-525. doi:10.1177/1073191110388146

Anagnostopoulos, F., Niakas, D., & Tountas, Y. (2009). Comparison between exploratory

   factor-analytic and SEM-based approaches to constructing SF-36 summary scores.

   Quality of Life Research, 18, 53-63. doi:10.1007/s11136-008-9423-5

Andrich, D., & van Schoubroeck, L. (1989). The General Health Questionnaire: A

   psychometric analysis using latent trait theory. *Psychological Medicine, 19*, 469-485.

   doi:10.1017/S0033291700012502

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for

   arbitrary rotation criteria in factor analysis. *Educational and Psychological

   Measurement, 65*, 676-696. doi:10.1177/0013164404272507

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic

   introduction to fixed-effect and random-effects models for meta-analysis. *Research

   Synthesis Methods, 1*, 97-111. doi:10.1002/jrsm.12

Browne, M. W. (1972). Orthogonal rotation to a partially specified target. *British Journal of

   Mathematical and Statistical Psychology*, 25, 115-120. doi:10.1111/j.2044-

   8317.1972.tb00482.x

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230-258. doi:10.1177/0049124192021002005

Campbell, A., Walker, J., & Farrell, G. (2003). Confirmatory factor analysis of the GHQ-12: Can I see that again? *Australian and New Zealand Journal of Psychiatry*, *37*, 475-483. doi:10.1046/j.1440-1614.2003.01208.x

Chin, E. G., Drescher, C. F., Trent, L. R., Darden, M., Seak, W. C., & Johnson, L. R. (2015). Searching for a screener: Examination of the factor structure of the General Health Questionnaire in Malaysia. *International Perspectives in Psychology: Research, Practice, Consultation*, *4*, 111-127. doi:10.1037/ipp0000030

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. doi:10.1177/001316446002000104

Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling, 20*, 429-454. doi:10.1080/10705511.2013.797827

Cheung, M. W. L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01521

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*, 40-64. doi:10.1037/1082-989X.10.1.40

Cheung, M. W. L., & Hong, R. Y. (2017). Applications of meta-analytic structural equation modeling in health psychology: Examples, issues, and recommendations. *Health Psychology Review, 11*, 265-279. doi:10.1080/17437199.2017.1343678

Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: a split/analyze/meta-analyze approach. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.00738

Debray, T., Moons, K. G., Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H., & Reitsma, J. B. (2015). Get real in individual participant data (PID) meta-analysis: a

review of the methodology. *Research Synthesis Methods*, *6*, 293-309.

doi:10.1002/jrsm.1160

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in g-factor models:

Explanations and alternatives. *Psychological Methods, 22*, 541-562.

doi:10.1037/met0000083

Fernandes, H. M., & Vasconcelos-Raposo, J. (2012). Factorial validity and invariance of the

GHQ-12 among clinical and nonclinical samples. *Assessment*, *20*, 219-229.

doi:10.1177/1073191112465768

Fryers, T., Brugha, T., Morgan, Z., Smith, J., Hill, T., Carta, M., ... & Kovess, V. (2004).

Prevalence of psychiatric disorder in Europe: the potential and reality of meta-

analysis. *Social Psychiatry and Psychiatric Epidemiology*, *39*, 899-905.

doi:10.1007/s00127-004-0875-9

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S.-C., & Cheung, Y.-B. (2004). Does the 12-

item General Health Questionnaire contain multiple factors and do we need them?

Health and Quality of Life Outcomes, 2:63. doi:10.1186/1477-7525-2-63

Gao, W., Stark, D., Bennett, M. I., Siegert, R. J., Murray, S., & Higginson, I. J. (2012). Using

the 12-item General Health Questionnaire to screen psychological distress from

survivorship to end-of-life care: Dimensionality and item quality. *Psycho-Oncology*,

21, 954-961. doi:10.1002/pon.1989

Gebhardt, F. (1968). Über die Ähnlichkeit von Faktormatrizen [On the similarity of factor

matrices]. *Psychologische Beiträge, 10*, 591-599.

Gnambs, T., & Kaspar, K. (2017). Socially desirable responding in web-based questionnaires:

A meta-analytic review of the candor hypothesis. *Assessment, 24*, 746-762.

doi:10.1177/1073191115624547

Gnambs, T., & Schroeders, U. (2017). Cognitive abilities explain wording effects in the

    Rosenberg Self-Esteem Scale. *Assessment*. Advance online publication.

    doi:10.1177/1073191117746503

Gnambs, T., & Staufenbiel, T. (2016). Parameter accuracy in meta-analyses of factor

    structures. *Research Synthesis Methods, 7*, 168-186. doi:10.1002/jrsm.1190

Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire*. London,

    England: Oxford University Press.

Goldberg, D. P, Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C.

    (1997). The validity of two versions of the GHQ in the WHO study of mental illness

    in general health care. *Psychological Medicine*, *27*, 191-197.

    doi:10.1017/S0033291796004242

Goldberg, D. P., & Williams, P. (1988*). A users's guide to the General Health Questionnaire*.

    London, England: GL Assessment.

Gouveia, V. V., Barbosa, G. A., Andrade, E. O., & Carneiro, M. B. (2010). Factorial validity

    and reliability of the General Health Questionnaire (GHQ-12) in the Brazilian physician

    population. *Cadernos de Saúde Pública, 26*, 1439-1445. doi:10.1590/S0102-

    311X2010000700023

Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social*

    Psychiatry and Psychiatric Epidemiology, 26, 132-138. doi:10.1007/BF00782952

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about

    model fit. *European Journal of Psychological Assessment, 33*, 313-317.

    doi:10.1027/1015-5759/a000450

Hankins, M. (2008a). The factor structure of the twelve item General Health Questionnaire

    (GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in*

    Mental Health, 4(10). doi:10.1186/1745-0179-4-10

Hankins, M. (2008b). The reliability of the twelve-item general health questionnaire (GHQ-

    12) under realistic assumptions. *BMC Public Health, 8*(355). doi:10.1186/1471-2458-8-

    355.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety

    Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical

    sample. *British Journal of Clinical Psychology, 44*, 227-239.

    doi:10.1348/014466505X29657

Horn, J. (1965). A rationale and test for the number of factors in factor analysis.

    Psychometrika, 30, 179-185. doi:10.1007/BF02289447

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an

    insidious confound in survey data. *Journal of Applied Psychology, 100*, 828-845.

    doi:10.1037/a0038510

Iwata, N., Okuyama, Y., Kawakami, Y., & Saito, K. (1988). The twelve-item General Health

    Questionnaire among Japanese workers. *Environmental Science, 11*, 1-10.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and

    response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*,

    264-277. doi:10.1177/0022022104272905

Kaiser, H. F. (1960). The application of the electronic computers to factor analysis.

    Educational and Psychological Measurement, 20, 141-151.

    doi:10.1177/001316446002000116

Khojasteh, K., & Lo, W.-J. (2015). Investigating the sensitivity of goodness-of-fit indices to

    detect measurement invariance in a bifactor model. *Structural Equation Modeling, 22*,

    531-541. doi:10.1080/10705511.2014.937791

Klainin-Yobas, P., & He, H. G. (2014). Testing psychometric properties of the 30-item

    general health questionnaire. *Western Journal of Nursing Research, 36*, 117-134.

    doi:10.1177/0193945913485649

Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous

    variables. *Psychology Science, 45*, 106-110.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability

    and interrater agreement. *Organizational Research Methods, 11*, 815-852.

    doi:10.1177/1094428106296642

Li, W. H. C., Chung, J. O. K., Chui, M. M. L., & Chan, P. S. L. (2009). Factorial structure of

    the Chinese version of the 12-item General Health Questionnaire in adolescents.

    Journal of Clinical Nursing, 18, 3253-3261. doi:10.1111/j.1365-2702.2009.02905.x

López-Castedo, A., & Fernández, L. (2005). Psychometric properties of the Spanish version

    of the 12-item General Health Questionnaire in adolescents. *Perceptual and Motor Skills,*

    100, 676-680. doi:10.2466/pms.100.3.676-680

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a

    meaningful index of factor similarity. *Methodology, 2*, 57-64. doi:10.1027/1614-

    2241.2.2.57

Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M. L., & Pulkkinen, L.

    (2006). The factor structure and factorial invariance of the 12-item General Health

    Questionnaire (GHQ-12) across time: Evidence from two community-based samples.

    Psychological Assessment, 18, 444-451. doi:10.1037/1040-3590.18.4.444

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful

    distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810-819.

    doi:10.1037/0022-3514.70.4.810

Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health

    Questionnaire. *Psychological Reports, 84*, 927-935. doi:10.2466/pr0.1999.84.3.927

Martin, C. R., & Newell, R. J. (2005). Is the 12-item General Health Questionnaire (GHQ-12)

    confounded by scoring method in individuals with facial disfigurement? *Psychology*

    and Health, *20*, 651-659. doi:10.1080/14768320500060061

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative

    fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568-

    592. doi:10.1037/0021-9010.93.3.568

Montazeri, A., Harirchi, A. M., Shariati, M., Garmaroudi, G., Ebadi, M., & Fateh, A. (2003).

    The 12-item General Health Questionnaire (GHQ-12): Translation and validation

    study of the Iranian version. *Health and Quality of Life Outcomes*, 1:66.

    doi:10.1186/1477-7525-1-66

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., …

    Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*, 1420-1422.

    doi:10.1126/science.aab2374

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science.

    Science, 349(6251), aac4716. doi:10.1126/science.aac4716

Piccinelli, M., Bisoffi, G., Bon, M. G., Cunico, L., & Tansella, M. (1993). Validity and test-

    retest reliability of the Italian version of the 12-item General Health Questionnaire in

    general practice: A comparison between three scoring methods. *Comprehensive*

    Psychiatry, *34*, 198-205. doi:10.1016/0010-440X(93)90048-9

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure

    of the 12-item General Health Questionnaire among young males in Italy. *Acta*

    Psychiatrica Scandinavica, *90*, 432-437. doi:10.1111/j.1600-0447.1994.tb01620.x

Quek, K. F., Low, W. Y., Razack, A. H., & Loh, C. S. (2001). Reliability and validity of the

    General Health Questionnaire (GHQ-12) among urological patients: A Malaysian

    study. *Psychiatry and Clinical Neurosciences, 55*, 509-513. doi:10.1046/j.1440-

    1819.2001.00897.x

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate*

    Behavioral Research, 47, 667-696. doi:10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations:

    Exploring the extent to which multidimensional data yield univocal scale scores. *Journal*

    of Personality Assessment, *92*, 544-559. doi:10.1080/00223891.2010.496477

Revelle, W. (2017). *psych: Procedures for personality and psychological research.*

    Northwestern University, Evanston, IL. Retrieved from http://CRAN.R-

    project.org/package=psych

Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of

    ambiguous response categories on the factor structure of the GHQ-12. *Psychological*

    Assessment, *26*, 1021-1030. doi:10.1037/a0036468

Romppel, M., Hinz, A., Finck, C., Young, J., Brähler, E., & Glaesmer, H. (2017). Cross-

    cultural measurement invariance of the General Health Questionnaire-12 in a German

    and a Colombian population sample. *International Journal of Methods in Psychiatric*

    Research. Advance online publication. doi:10.1002/mpr.1532

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models:

    Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137-150.

    doi:10.1037/met0000045

Saris, W. E., Satorra, A. & van der Veld, W. M. (2009). Testing structural equation models or

    detection of misspecifications. *Structural Equation Modeling, 16*, 561-582.

    doi:10.1080/10705510903203433

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural

    equation models: Tests of significance and descriptive goodness-of-fit measures.

    Methods of Psychological Research Online, 8, 23-74.

Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health

    Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatrica

    Scandinavica*, *100*, 462-468. doi:10.1111/j.1600-0447.1999.tb10898.x

Schulze, R. (2005). Modeling structures of intelligence. In O. Wilhelm & R. W. Engle (Eds.),

    Handbook of understanding and measuring intelligence (pp. 241-263). Thousand

    Oaks, CA: Sage Publications.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the

    GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychological

    Assessment*, *17*, 231-236. doi:10.1037/1040-3590.17.2.231

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability.

    Psychological Bulletin, 86, 420-428. doi:10.1037/0033-2909.86.2.420

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor

    structure of the GHQ-12: The interaction between item phrasing, variance and levels

    of distress. *Quality of Life Research, 22*, 145-152. doi:10.1007/s11136-012-0133-7

Tait, R. J., French, D. J., & Hulse, G. K. (2003). Validity and psychometric properties of the

General Health Questionnaire-12 in young Australian adolescents. *Australian and New

Zealand Journal of Psychiatry*, *37*, 374-381. doi:10.1046/j.1440-1614.2003.01133.x

Tomás, J. M., Gutiérrez, M., & Sancho, P. (2017). Factorial validity of the General Health

Questionnaire 12 in an Angolan sample. *European Journal of Psychological

Assessment, 33*, 116-112. doi:10.1027/1015-5759/a000278

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research

Section Report No. 984). Washington, DC: Department of the Army.

Vanheule, S., & Bogaerts, S. (2005). The factorial structure of the GHQ-12. *Stress and

Health*, *21*, 217-222. doi:10.1002/smi.1058

Vasconcelos-Raposo, J., Fernandes, H. M., & Teixeira, C. M. (2013). Factor structure and

reliability of the depression, anxiety and stress scales in a large Portuguese community

sample. *Spanish Journal of Psychology, 16*, 1-10. doi:10.1017/sjp.2013.15

Velicer, W. (1976). Determining the number of components from the matrix of partial

correlations. *Psychometrika, 41*, 321-327. doi:10.1007/BF02293557

Wang, L., & Lin, W. (2011). Wording effects and the dimensionality of the General Health

Questionnaire (GHQ-12). *Personality and Individual Differences*, *50*, 1056-1061.

doi:10.1016/j.paid.2011.01.024

Wong, K. C. K., & O'Driscoll, M. P. (2016). Psychometric properties of the General Health

Questionnaire-12 in a sample of Hong Kong employees. *Psychology, Health &

Medicine*. doi:10.1080/13548506.2016.1140901

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of

wording effects. *Personality and Individual Differences*, *46*, 197-201.

doi:10.1016/j.paid.2008.09.027

Footnotes

1)   Random-effects meta-analyses using a small number of samples can result in unstable

estimates of between-studies variances (Borenstein, Hedges, Higgins, & Rothstein,

2010). Accordingly, for some subgroup analyses respective random-effects model did

not converge and did not give meaningful heterogeneity estimates for several pooled

correlations. Therefore, subgroup analyses pertaining to correlation matrices as effects

sizes as well as different language versions were based on a fixed-effects model,

whereas all other analyses adopted a random-effects model.

Table 1.

Exploratory Factor Loading Patterns in Meta-Analysis I

		Single factor model		Oblique factor model			Bifactor model				
	Item	Factor 1	h^2	Factor 1	Factor 2	h^2	General factor	Specific factor 1	Specific factor 2	h^2	I-ECV
I01	Been able to concentrate on whatever you are doing	**.51**	.26	.17	**.41**	.28	**.49**	.12	.25	.32	.78
I03	Felt that you are playing a useful part in things	**.49**	.24	-.04	**.63**	.37	**.47**	-.11	.37	.37	.59
I04	Felt capable of making decisions about things	**.49**	.24	-.06	**.65**	.38	**.46**	-.07	.38	.37	.58
I07	Enjoyed normal day-to-day activities	**.54**	.29	.13	**.49**	.34	**.52**	.07	.29	.36	.76
I08	Been able to face up to your problems	**.52**	.27	.02	**.60**	.37	**.50**	-.10	.33	.36	.67
I12	Felt reasonably happy, all things considered	**.51**	.26	.09	**.50**	.32	**.49**	-.02	.28	.32	.75
I02	Lost much sleep over worry	**.54**	.29	**.65**	-.07	.37	**.56**	.27	-.17	.41	.78
I05	Felt constantly under strain	**.61**	.37	**.73**	-.07	.48	**.63**	.25	-.19	.50	.79
I06	Felt you could not overcome your difficulties	**.61**	.37	**.66**	.01	.44	**.65**	.06	-.17	.44	.94
I09	Been feeling unhappy and depressed	**.68**	.47	**.72**	.04	.55	**.72**	.08	-.16	.54	.94
I10	Been losing confidence in yourself	**.66**	.43	**.62**	.11	.47	**.70**	-.16	-.16	.53	.87
I11	Thinking of yourself as a worthless person	**.57**	.33	**.50**	.13	.34	**.60**	-.23	-.13	.44	.85
	Eigenvalue	3.81		2.71	2.01		3.92	0.27	0.78		
	Proportion of variance	32%		23%	17%		33%	2%	6%		
	Proportion of explained variance	100%		57%	43%		79%	5%	16%		

Note. Exploratory weighted least square factor analysis with direct oblimin (Bernaards & Jennrich, 2005) or target rotation (Browne, 1972). The factor correlation in the oblique case was r = .61. All items were recoded in such a way that higher values indicate better mental health. Factor loadings $|\lambda| \geq .40$ are in bold. h^2 = Communality; I-ECV = Proportion of common variance explained by the general factor (Rodriguez et al., 2016).

Table 2.

Fit Statistics for Different Confirmatory Factor Models of the GHQ-12.

	Model				*Meta-Analysis I*								*Meta-Analysis II*				
		χ^2	df	CFI	SRMR	RMSEA	AIC	BIC	χ^2	df	CFI	SRMR	RMSEA	AIC	BIC
1.	Single factor model	1,007*	54	.939	.066	.015	899	400	159,642*	54	.894	.145	.084	159,534	158,944
2.	Artifactual model	173*	48	.992	.027	.006	77	-366	96,328*	48	.936	.086	.070	96,232	95,707
Andrich & van Schoubroeck (1989)																
3a.	Correlated factor model	207*	53	.990	.029	.006	101	-389	98,620*	53	.935	.088	.067	98,514	97,935
3b.	Bifactor model	75*	42	.998	.016	.003	-9	-397	43,575*	42	.971	.040	.050	43,491	43,032
Graetz (1991)																
4a.	Correlated factor model	157*	51	.993	.025	.005	55	-416	78,832*	51	.948	.072	.061	78,730	78,173
4b.	Bifactor model	75*	43	.998	.016	.003	-11	-408	43,705*	43	.971	.040	.050	43,619	43,149
Martin (1999)																
5a.	Correlated factor model	716*	51	.958	.056	.013	614	142	136,571*	51	.910	.123	.081	136,469	135,912
5b.	Bifactor model	494*	42	.971	.046	.012	410	22				Model did not converge			

Note. N = 76,473 and 410,640 for meta-analyses I and II. CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Square Error of

Approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

* $p < .05$

Table 3.

Score Reliabilities for Different General Factor Models of the GHQ-12.

Model	ω_h	ω_{s_1}	ω_{s_2}	ω_{s_3}	ECV_g	ECV_{s_1}	ECV_{s_2}	ECV_{s_3}
Meta-Analysis I								
1. Single factor model	.89							
2. Artifactual model	.75	.47			.70	.30		
3. Andrich & van Schoubroeck (1989)	.78	.43	.03		.72	.22	.06	
4. Graetz (1991)	.77	.41	.15	.08	.69	.21	.07	.02
5. Martin (1999)	.86	.31	.06	.06	.75	.12	.05	.08
Meta-Analysis II								
1. Single factor model	.96							
2. Artifactual model	.85	.36			.79	.21		
3. Andrich & van Schoubroeck (1989)	.85	.39	.01		.77	.18	.05	
4. Graetz (1991)	.84	.35	.15	.13	.74	.16	.07	.03

Note. ω_h = General factor reliability (i.e., proportion of variance in total scores attributed to the general factor); ω_s = Specific factor reliability (i.e., proportion of variance in subscale scores attributed to the specific factor); ECV = Proportion of common variance explained by the general / specific factor (see Rodriguez et al., 2016); s_1 = Items 1, 3, 4, 7, 8, and 12 (Models 2 and 3), 2, 5, 6, and 9 (Model 4), or 6, 9, 10, 11, and 12 (Model 5); s_2 = Items 2, 5, 6, 9, 10, and 11 (Model 3), 1, 3, 4, 7, 8, and 12 (Model 4), or 2, 5, and 7 (Model 5); s_3 = Items 10 and 11 (Model 4) or 1, 3, 4, and 8 (Model 5).

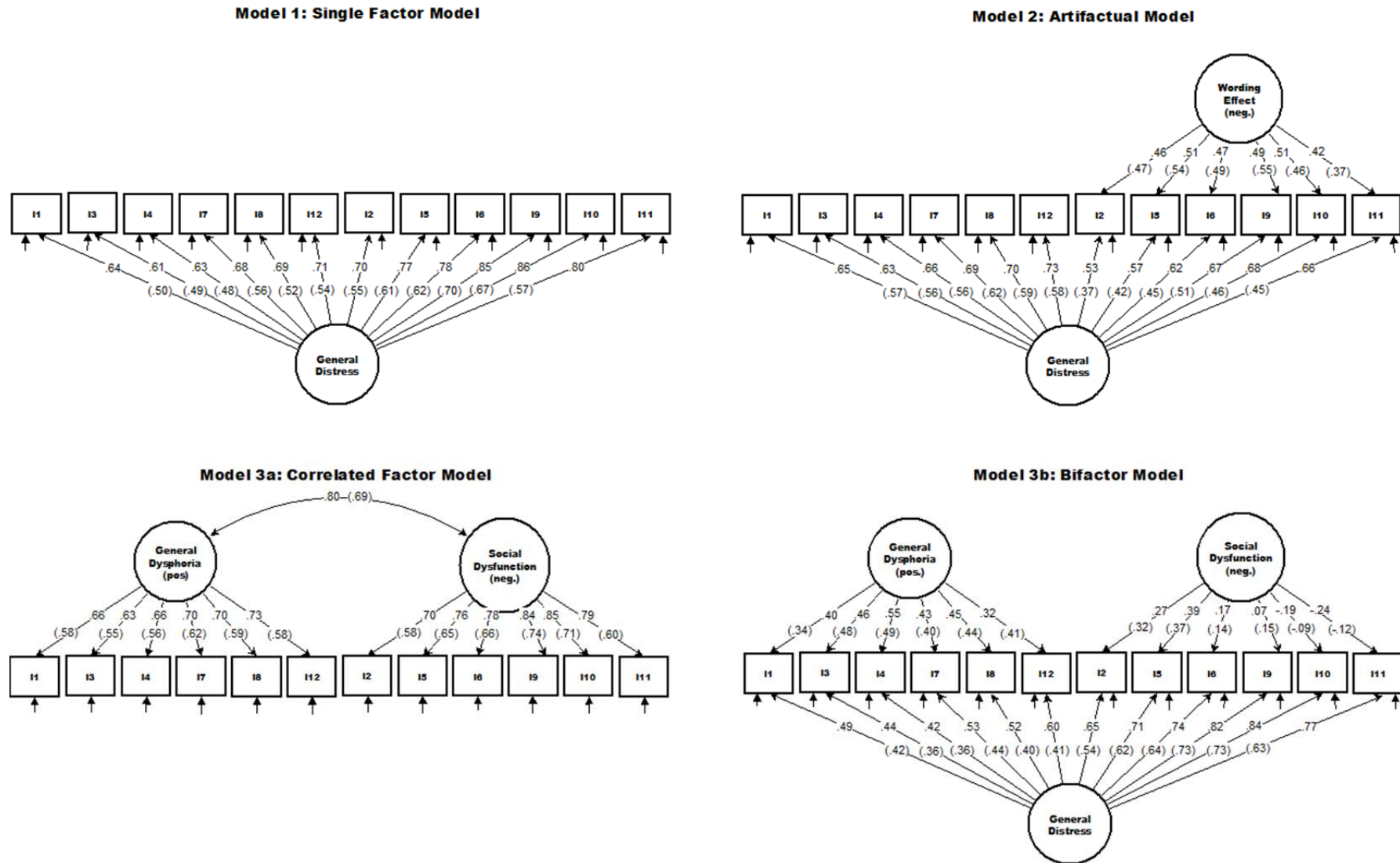*Figure 1.* Factor models for the GHQ-12 with standardized parameter estimates for Meta-analysis I (in parentheses) and II.

Supplemental Material for

"The Structure of the General Health Questionnaire (GHQ-12):

Two Meta-Analytic Factor Analyses"

Timo Gnambs & Thomas Staufenbiel

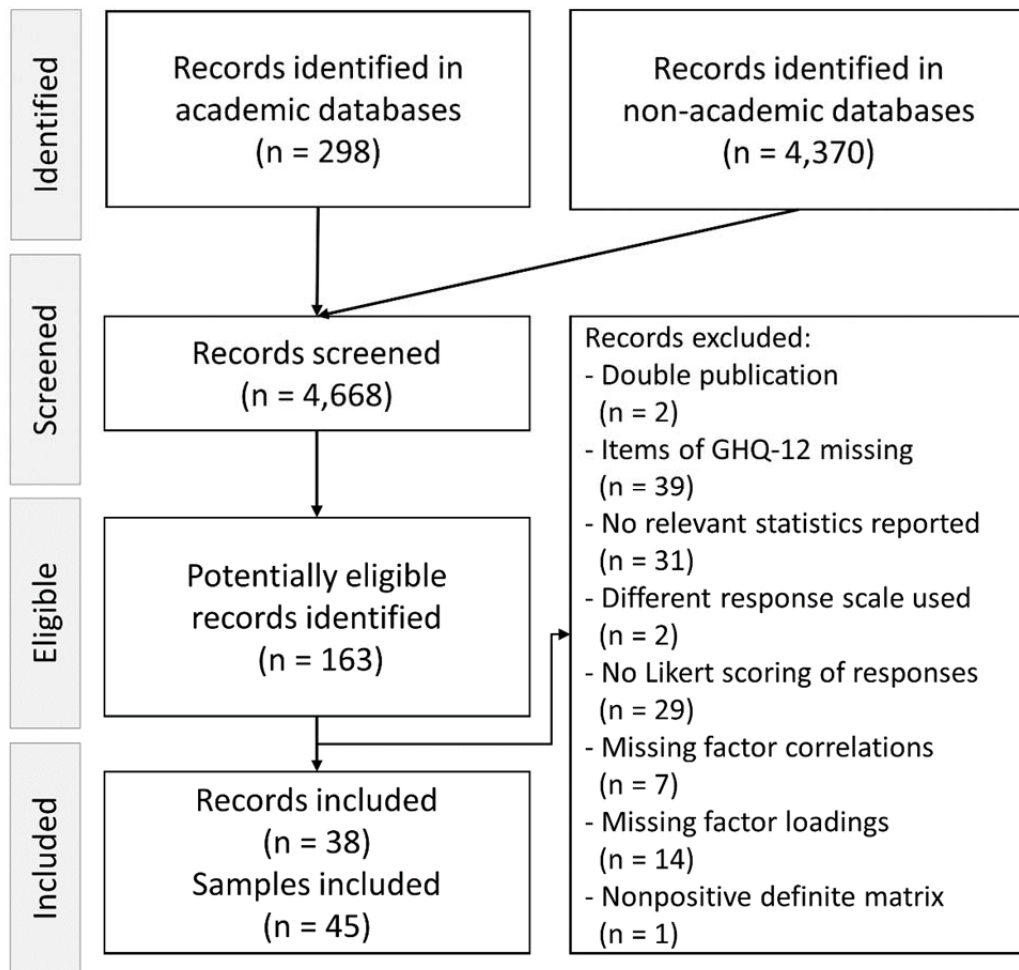**Flow Chart of Search Process for Meta-Analysis I**

Figure S1. Flow chart of search process (Meta-analysis I)

Pooled Correlation Matrix for the GHQ-12 in Meta-Analysis I

	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12
Pooled correlations (lower off diagonal) and standard errors (upper off diagonal)												
I01		.016	.015	.014	.015	.014	.017	.016	.013	.013	.011	.018
I02	.268		.015	.017	.016	.015	.017	.017	.016	.016	.016	.018
I03	.310	.175		.018	.013	.014	.014	.017	.016	.015	.016	.013
I04	.320	.187	.395		.016	.015	.017	.020	.013	.015	.014	.017
I05	.297	.458	.198	.216		.016	.019	.015	.020	.016	.015	.020
I06	.271	.392	.222	.229	.454		.015	.018	.017	.015	.016	.018
I07	.326	.247	.338	.323	.289	.293		.017	.017	.016	.015	.015
I08	.307	.191	.346	.372	.226	.265	.355		.016	.017	.014	.017
I09	.314	.445	.268	.248	.508	.477	.343	.293		.015	.015	.024
I10	.286	.370	.284	.276	.426	.464	.307	.303	.522		.020	.018
I11	.241	.306	.261	.244	.353	.486	.249	.282	.432	.489		.020
I12	.294	.224	.326	.315	.243	.262	.372	.344	.317	.294	.271	
Random variances (lower off diagonal) and I^2 (upper off diagonal)												
I01		.955	.944	.934	.948	.930	.960	.952	.929	.924	.893	.966
I02	.010		.936	.951	.972	.956	.955	.956	.967	.962	.953	.960
I03	.008	.008		.970	.919	.927	.939	.962	.951	.943	.949	.933
I04	.006	.011	.012		.943	.942	.964	.974	.918	.947	.930	.960
I05	.008	.010	.006	.009		.968	.968	.942	.985	.966	.949	.970
I06	.006	.008	.007	.009	.010		.948	.965	.975	.964	.961	.961
I07	.011	.011	.007	.011	.014	.009		.963	.961	.954	.940	.950
I08	.009	.011	.011	.015	.008	.013	.011		.956	.962	.940	.965
I09	.006	.009	.009	.006	.016	.012	.011	.010		.972	.962	.981
I10	.006	.009	.008	.009	.010	.009	.010	.011	.009		.983	.967
I11	.004	.009	.009	.007	.008	.010	.008	.007	.009	.016		.970
I12	.013	.012	.006	.011	.016	.012	.008	.012	.023	.013	.015	

Note. N = 76,473.

Pooled Correlation Matrices for the GHQ-12 in Meta-Analysis II

	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12
I01		.009	.007	.006	.006	.008	.006	.011	.005	.009	.008	.048
I02	.339		.007	.021	.022	.016	.006	.019	.028	.009	.005	.007
I03	.351	.222		.010	.012	.007	.006	.040	.011	.007	.027	.009
I04	.398	.224	.454		.009	.015	.009	.007	.013	.010	.012	.011
I05	.349	.557	.216	.234		.009	.019	.011	.009	.015	.005	.014
I06	.351	.487	.305	.310	.579		.021	.033	.072	.007	.041	.005
I07	.431	.330	.379	.349	.379	.382		.009	.033	.017	.009	.005
I08	.388	.300	.361	.451	.323	.387	.451		.012	.008	.010	.007
I09	.377	.546	.301	.285	.580	.582	.409	.392		.017	.012	.006
I10	.372	.464	.348	.338	.489	.562	.377	.402	.656		.050	.005
I11	.324	.397	.354	.321	.409	.504	.337	.486	.566	.676		.005
I12	.370	.350	.359	.368	.375	.400	.449	.453	.499	.451	.451	

Note. $N = 410{,}640$. Values in the lower off diagonal are pooled correlations and values in the upper off diagonal are standard errors.

Studies Excluded From Meta-Analysis I

Study	Reason for exclusion
Abubakar & Fischer (2012)	No relevant statistics reported
Aderibigbe et al. (1996)	Items of GHQ-12 missing
Aguado et al. (2012)	No relevant statistics reported
Ang (2011)	No relevant statistics reported
van Baalen (2013)	Items of GHQ-12 missing
Bakhla et al. (2013)	Different response format used
Berwick et al. (1987)	Missing factor loadings
Brabete (2014)	No Likert scoring
Çebi (2009)	No Likert scoring
Campo-Arias (2007)	Missing factor correlations
Caraveo-Anduaga et al. (1998)	No Likert scoring
Carvalho et al. (2010)	No relevant statistics reported
Castro-Costa et al. (2008)	No relevant statistics reported
Chan (1985)	Missing factor loadings
Chan (1993)	Items of GHQ-12 missing
Chan (1995)	Items of GHQ-12 missing
Cheung (2002)	No relevant statistics reported
Chin et al. (2015)	No relevant statistics reported
Cleary et al. (1982)	Unknown response scoring
D'Arcy (1982)	No Likert scoring
Dale et al. (2012)	Missing factor loadings
Daradkeh et al. (2012)	Missing factor loadings
Delgado-Gomez et al. (2013)	Items of GHQ-12 missing
Doi & Minowa (2003)	No Likert scoring
Elton et al. (1988)	Items of GHQ-12 missing
Epstein et al. (1994)	Unknown response scoring
Failde et al. (2000)	Items of GHQ-12 missing
Farrell (1998)	No Likert scoring
Fernandes & Vasconcelos-Raposo (2012)	No relevant statistics reported
French & Tait (2004)	No relevant statistics reported
Gallego & Soria (2000)	No relevant statistics reported
Gao et al. (2004)	No relevant statistics reported

Study	Reason for exclusion
Gao et al. (2012)	No Likert scoring
Gelaye et al. (2015)	No Likert scoring
Ghanbarnejad et al. (2013)	Items of GHQ-12 missing
Gibbons et al. (2004)	Items of GHQ-12 missing
Goldberg & Hillier (1979)	Items of GHQ-12 missing
Gouveia et al., (2010)	Nonpositive definite correlation matrix
Hobi et al. (1989)	Items of GHQ-12 missing
Horton (2011)	Missing factor correlations
Ibrahim et al. (2014)	Missing factor loadings
Ignatyev et al. (2012)	Items of GHQ-12 missing
Ip & Martin (2006)	No relevant statistics reported
Iwata & Saito (1992)	Items of GHQ-12 missing
Jain & Sinha (2005)	Different response format used
Kalliath et al. (2004)	No relevant statistics reported
Klainin-Yobas & He (2014)	No relevant statistics reported
de Kock et al. (2014)	Items of GHQ-12 missing
Krol et al (1994)	Items of GHQ-12 missing
Laranjeira (2008)	No Likert scoring
Layton & Rust (1986)	No Likert scoring
Li et al. (2009)	Missing factor correlations
Liang et al. (2016)	Missing factor loadings
López-Castedo & Domínguez (2010)	Items of GHQ-12 missing
Mäkikangas et al. (2007)	No relevant statistics reported
Malakouti et al. (2007)	Items of GHQ-12 missing
Mansor et al. (2016)	Unknown response scoring
Martin & Newell (2005)	No relevant statistics reported
Martin (1999)	No relevant statistics reported
Medina-Mora et al. (1983)	Items of GHQ-12 missing
Meertens (2004)	Items of GHQ-12 missing
Molavi (2002)	Items of GHQ-12 missing
Molina & Andrade (2002)	No relevant statistics reported
Molina et al. (2006)	Items of GHQ-12 missing
Molina et al. (2014)	No relevant statistics reported

Study	Reason for exclusion
Montazeri et al. (2003)	No Likert scoring
Monteiro (2011)	Items of GHQ-12 missing
Munyombwe et al. (2015)	Items of GHQ-12 missing
Nagyova et al. (2000)	Items of GHQ-12 missing
Najarkolaei et al. (2014)	No Likert scoring
Netz et al. (2012)	Items of GHQ-12 missing
O'Connor & Parslow (2010)	Missing factor correlations
Ohta et al. (1995)	Missing factor loadings
Ohta et al. (1998)	Missing factor loadings
Ohta et al. (2003)	Identical sample to Ohta (1998)
Ozaki et al. (2002)	No Likert scoring
Padron et al. (2012)	Missing factor correlations
Pariente et al. (1992)	Items of GHQ-12 missing
Parkes (1982)	Items of GHQ-12 missing
Petkovska et al. (2015)	Missing factor loadings
Picardi et al. (2001)	Missing factor loadings
Politi et al. (1994)	Identical sample to Piccinelli & Politi (1993)
Prady et al. (2013)	Items of GHQ-12 missing
Rajabi & Sheykhshabani (2009)	Items of GHQ-12 missing
Rivas-Diez & Sánchez-López (2014)	Missing factor loadings
Rocha et al. (2011)	Missing factor correlations
Romppel et al. (2013)	No relevant statistics reported
Salama-Younes et al. (2009)	No relevant statistics reported
Sánchez-Lopéz & Dresch (2008)	Missing factor loadings
Shek (1987)	Unknown response scoring
Shek (1993)	No relevant statistics reported
Shek & Tsang (1995)	Unknown response scoring
Shevlin & Adamson (2005)	No relevant statistics reported
Shigemi et al. (2000)	Missing factor loadings
Siegert & Chung (1995)	Items of GHQ-12 missing
Siegert et al. (1987)	Items of GHQ-12 missing
Smith et al. (2013)	No relevant statistics reported
Stevenson (1989)	No Likert scoring

Study	Reason for exclusion
Stuart et al. (1993)	Items of GHQ-12 missing
Sun et al. (2012)	No Likert scoring
Suzuki et al. (2011)	No Likert scoring
Sveen et al. (2004)	Items of GHQ-12 missing
Takeuchi & Kitamura (1991)	Items of GHQ-12 missing
Talwar & Rahman (2015)	No Likert scoring
Thabet (2005)	Items of GHQ-12 missing
Tomás et al. (2015)	No relevant statistics reported
Toyabe et al. (2006)	No relevant statistics reported
Urzúa et al. (2015)	No relevant statistics reported
Vallejo et al. (2007)	Items of GHQ-12 missing
Vanheule & Bogaerts (2005)	Items of GHQ-12 missing
Vázquez-Barquero et al. (1988)	Missing factor loadings
Villa et al. (2013)	No relevant statistics reported
Wang & Lin (2011)	No relevant statistics reported
Weyerer et al. (1986)	Items of GHQ-12 missing
Winzer et al. (2014)	Missing factor correlations
Wong & O`Driscoll (2015)	No relevant statistics reported
Ye (2009)	No relevant statistics reported
Zelča et al. (2013)	No Likert scoring
Zulkefly et al. (2010)	No Likert scoring

References for excluded studies

Abubakar, A., & Fischer, R. (2012). The factor structure of the 12-item General Health

Questionnaire in a literate Kenyan population. *Stress and Health, 28*, 248-254.

doi:10.1002/smi.1420

Aderibigbe, Y. A., Riley, W., Lewin, T., & Gureje, O. (1996). Factor structure of the 28-item

General Health Questionnaire in a sample of antenatal women. *International Journal

of Psychiatry in Medicine, 26*, 263-269. doi:10.2190/3XAV-M1BC-DA2B-DCMF

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., & Luciano, J. V.

(2012). Examining the factor structure and discriminant validity of the 12-item

General Health Questionnaire (GHQ-12) among Spanish postpartum women.

Assessment, 19, 517-525. doi:10.1177/1073191110388146

Ang, H. B. (2011). *Core self-evaluations as a moderator of the stressor-strain-satisfaction

relationships among farm managers in New Zealand* (Doctoral thesis, University of

Auckland, Australia). Retrieved from http://hdl.handle.net/2292/6834

van Baalen, S. (2013). *Generating insights into the elusive concept of overqualification*

(Master's thesis, Tilburg University, The Netherlands). Retrieved from

http://arno.uvt.nl/show.cgi?fid=130425

Bakhla, A. K., Verma, V., Hembram, M., Praharaj, S. K., & Sinha, V. K. (2013). Internal

consistency and factor structure of 12-item General Health Questionnaire in visually

impaired students. *Industrial Psychiatry Journal, 22*, 109-113. doi:10.4103/0972-

6748.132918

Berwick, D. M., Budman, S., Damico-White, J., Feldstein, M., & Klerman, G. L. (1987).

Assessment of psychological morbidity in primary care: explorations with the General

Health Questionnaire. *Journal of Chronic Diseases, 40*, 71S-79S. doi:10.1016/S0021-9681(87)80035-8

Brabete, A. C. (2014). El Cuestionario de Salud General de 12 items (GHQ-12): estudio de traducción y adaptación de la version rumana [The 12-Item General Health Questionnaire (GHQ-12): a translation and adaptation study of the Romanian version]. *Revista Iberoamericana de Diagnóstico y Evaluación, 37*, 9-27.

Campo-Arias, A. (2007). Cuestionario General de Salud-12: análisis de factores en población general de Bucaramanga, Colombia [General Health Questionnaire-12: factor analysis in a general population of Bucaramanga, Colombia]. *Iatreia, 20*, 29-36.

Caraveo-Anduaga, J. J., Martínez, N. A., Saldívar, G., López, J. L., & Saltijeral, M. T. (1998). Performance of the GHQ-12 in relation to current and lifetime CDI psychiatric diagnoses. *Salud Mental, 21*, 1-11.

Carvalho, H. W., Patrick, C. J., Jorge, M. R., & Andreoli, S. B. (2010). Validation of the structural coherency of the General Health Questionnaire. *Revista Brasileira de Psiquiatria, 33*, 59-63.

Castro-Costa, E. D. C., Uchoa, E., Firmo, J. O. A., Lima-Costa, M. F., & Prince, M. (2008). Association of cognitive impairment, activity limitation with latent traits in the GHQ-12 in the older elderly. The Bambui Health and Aging Study. *Aging Clinical and Experimental Research, 20*, 562-568. doi:10.1007/BF03324885

Çebi, E. (2009). *University students' attitudes toward seeking psychological help: Effects of perceived social support, psychological distress, prior help-seeking experience and gender* (Master's thesis, Middle East Technical University, Turkey). Retrieved from https://etd.lib.metu.edu.tr/upload/12610828/index.pdf

Chan, D. W. (1985). The Chinese version of the General Health Questionnaire: does language

make a difference? *Psychological Medicine, 15*, 147-155.

doi:10.1017/S0033291700021012

Chan, D. W. (1993). The Chinese General Health Questionnaire in a psychiatric setting: the

development of the Chinese scaled version. *Social Psychiatry and Psychiatric*

Epidemiology, 28, 124-129. doi:10.1007/BF00801742

Chan, D. W. (1995). The two scaled versions of the Chinese General Health Questionnaire: A

comparative analysis. *Social Psychiatry and Psychiatric Epidemiology*, *30*, 85-91.

doi:10.1007/BF00794948

Cheung, Y. B. (2002). A confirmatory factor analysis of the 12-item General Health

Questionnaire among older people. *International Journal of Geriatric Psychiatry, 17*,

739-744. doi:10.1002/gps.693

Chin, E. G., Drescher, C. F., Trent, L. R., Seak, W. C., & Johnson, L. R. (2015). Searching for

a screener: Examination of the factor structure of the General Health Questionnaire in

Malaysia. *International Perspectives in Psychology: Research, Practice, Consultation,*

4, 111-127. doi:10.1037/ipp0000030

Cleary, P. D., Goldberg, I. D., Kessler, L. G., & Nycz, G. R. (1982). Screening for mental

disorder among primary care patients: Usefulness of the General Health

Questionnaire. *Archives of general psychiatry, 39*, 837-840.

doi:10.1001/archpsyc.1982.04290070065012

Dale, B., Söderhamn, U., & Söderhamn, O. (2012). Psychometric properties of the Norwegian

version of the General Health Questionnaire (GHQ-30) among older people living at

home. *Psychology Research and Behavior Management, 5*, 151-157.

doi:10.2147/PRBM.S37502

Daradkeh, T. K., Ghubash, R., & El-Rufaie, O. E. F. (2001). Reliability, validity, and factor

structure of the Arabic version of the 12-item General Health Questionnaire.

Psychological Reports, 89, 85-94. doi:10.2466/pr0.2001.89.1.85

D'Arcy, C. (1982). Prevalence and correlates of nonpsychotic psychiatric symptoms in the

general population. *Canadian Journal of Psychiatry, 27*, 316-324.

Delgado-Gomez, D., Lopez-Castroman, J., de Leon-Martinez, V., Baca-Garcia, E., Cabanas-

Arrate, M. L., Sanchez-Gonzalez, A., & Aguado, D. (2013). Psychometrical

assessment and item analysis of the General Health Questionnaire in victims of

terrorism. *Psychological Assessment, 25*, 279-287. doi:10.1037/a0030645

Doi, Y., & Minowa, M. (2003). Factor structure of the 12-item General Health Questionnaire

in the Japanese general adult population. *Psychiatry and Clinical Neurosciences, 57*,

379-383.

Elton, M., Patton, G., Weyerer, S., Diallina, M., & Fichter, M. (1988). A comparative

investigation of the principal component structure of the 28 item version of the

General Health Questionnaire (GHQ). *Acta Psychiatrica Scandinavica, 77*, 124-132.

doi:10.1111/j.1600-0447.1988.tb05088.x

Epstein, R. S., Fullerton, C. S., & Ursano, R. J. (1994). Factor analysis of the General Health

Questionnaire. *Psychological Reports, 75*, 979-983. doi:10.2466/pr0.1994.75.2.979

Failde, I., Ramos, I., & Fernandez-Palacin, F. (2000). Comparison between the GHQ-28 and

SF-36 (MH 1–5) for the assessment of the mental health in patients with ischaemic

heart disease. *European Journal of Epidemiology, 16*, 311-316.

doi:10.1023/A:1007688525023

Farrell, G. A. (1998). The mental health of hospital nurses in Tasmania as measured by the

    12-item General Health Questionnaire. *Journal of Advanced Nursing, 28*, 707-712.

    doi:10.1046/j.1365-2648.1998.00735.x

Fernandes, H. M., & Vasconcelos-Raposo, J. (2012). Factorial validity and invariance of the

    GHQ-12 among clinical and nonclinical samples. *Assessment, 20*, 219-229.

    doi:10.1177/1073191112465768

French, D. J., & Tait, R. J. (2004). Measurement invariance in the General Health

    Questionnaire-12 in young Australian adolescents. *European Child & Adolescent*

    Psychiatry, 13, 1-7. doi:10.1007/s00787-004-0345-7

Gallego, C., & Soria, S. (2000). Validación factorial del General Health Questionnaire (GHQ-

    12) mediante un análisis factorial confirmatorio [Factor validation of the General

    Health Questionnaire (GHQ-12) with a confirmatory factor analysis]. *Revista de*

    Psicologia de la Salud, 12, 75-89.

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S.-C., & Cheung, Y.-B. (2004). Does the 12-

    item General Health Questionnaire contain multiple factors and do we need them?

    Health and Quality of Life Outcomes, 2:63. doi:10.1186/1477-7525-2-63

Gao, W., Stark, D., Bennett, M. I., Siegert, R. J., Murray, S., & Higginson, I. J. (2012). Using

    the 12-item General Health Questionnaire to screen psychological distress from

    survivorship to end-of-life care: dimensionality and item quality. *Psycho-Oncology,*

    21, 954-961. doi:10.1002/pon.1989

Gelaye, B., Tadesse, M. G., Lohsoonthorn, V., Lertmeharit, S., Pensuksan, W. C., Sanchez, S.

    E., ... & Anderade, A. (2015). 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. *Journal of Affective Disorders, 187*, 197-202. doi:10.1016/j.jad.2015.08.045

Ghanbarnejad, A., Turki, H., Golestan, B., Arbabi, M., Nedjat, S. (2013). Factor structure of Persian General Health Questionnaire-28 in dermatologic patients: A confirmatory factor analysis. *International Electronic Journal of Medicine, 2*, 11-21.

Gibbons, P., de Arévalo, H. F., & Mónico, M. (2004). Assessment of the factor structure and reliability of the 28 item version of the General Health Questionnaire (GHQ-28) in El Salvador. *International Journal of Clinical and Health Psychology, 4*, 389-398.

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9*, 139-145. doi:10.1017/S0033291700021644

Gouveia, V. V., Barbosa, G. A., Andrade, E. O., & Carneiro, M. B. (2010). Factorial validity and reliability of the General Health Questionnaire (GHQ-12) in the Brazilian physician population. *Cadernos de Saúde Pública, 26*, 1439-1445. doi:10.1590/S0102-311X2010000700023

Hobi, V., Gerhard, U., & Gutzwiller, F. (1989). A report on experiences using Goldberg's GHQ (General Health Questionnaire). *Schweizerische Rundschau für Medizin Praxis*, *78*, 219-225.

Horton, T. E. (2011). *Workaholism and employee well-being* (Master's thesis, University of Waikato, New Zealand). Retrieved from http://hdl.handle.net/10289/5758

Ibrahim, N., Osman, Z. J., Jan, K. O. N. N., Ismail, S. I. F., Kar, P. C., Mukhtar, F., & Sidik, S. M. (2014). Reliability and Factor structure of the General Health Questionnaire-12 among university students. *Malaysian Journal of Medicine and Health Sciences, 10*, 53-60.

Ignatyev, Y., Assimov, M., Aichberger, M. C., Ivens, S., Mir, J., Dochshanov, D., Ströhle, A., Heinz, A., & Mundt, A. P. (2012). Psychometric properties of a Russian version of the General Health Questionnaire-28. *Psychopathology, 45*, 252-258. doi:10.1159/000334918

Ip, W. Y., & Martin, C. R. (2006). Psychometric properties of the 12-item General Health Questionnaire (GHQ-12) in Chinese women during pregnancy and in the postnatal period. *Psychology, Health, & Medicine, 11*, 60-60. doi:10.1080/13548500500155750

Iwata, N., & Saito, K. (1992). The factor structure of the 28-Item General Health Questionnaire when used in Japanese early adolescents and adult employees: Age- and cross-cultural comparisons. *European Archives of Psychiatry and Clinical Neuroscience, 242*, 172-178. doi:10.1007/BF02191565

Jain, A. K., & Sinha, A. K. (2005). General health in organizations: Relative relevance of emotional intelligence, trust, and organizational support. *International Journal of Stress Management, 12*, 257-273. doi:10.1037/1072-5245.12.3.257

Kalliath, T. J., O'Driscoll, M. P., & Brough, P. (2004). A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health, 20*, 11-20. doi:10.1002/smi.993

Klainin-Yobas, P., & He, H.-G. (2014). Testing psychometric properties of the 30-item General Health Questionnaire. *Western Journal of Nursing Research, 36*, 117-134. doi:10.1177/0193945913485649

de Kock, F. S., Görgens-Ekermans, G., & Dhladhla, T. J. (2014). A confirmatory factor analysis of the General Health Questionnaire–28 in a Black South African sample. *Journal of Health Psychology, 19*, 1222-1231. doi:10.1177/1359105313488972

Krol, B., Sanderman, R., Moum, T., Suurmeijer, T., Doeglas, D., Krijnen, W., …, van den Heuvel, W. (1994). A comparison of the General Health Questionnaire-28 between

patients with rheumatoid arthritis from the Netherlands, France, Sweden and Norway. *European Journal of Psychological Assessment, 10*, 93-101.

Laranjeira, C. A. (2008). General Health Questionnaire-12 items: adaptation study to the Portuguese population. *Epidemiologia e Psichiatria Sociale, 17*, 148-151. doi:10.1017/S1121189X00002840

Layton, C., & Rust, J. (1986). The factor structure of the 60 item General Health Questionnaire. *Social Behavior and Personality, 14*, 123-131. doi:10.2224/sbp.1986.14.2.123

Liang, Y., Wang, L., & Yin, X. (2016). The factor structure of the 12-item general health questionnaire (GHQ-12) in young Chinese civil servants. *Health and Quality of Life Outcomes, 14*, 136. doi:10.1186/s12955-016-0539-y

López-Castedo, A., & Domínguez, J. (2010). Exploratory factor analysis and psychometric properties of the General Health Questionnaire in Spanish adolescents. *Psychological Reports, 107*, 120-126. doi:10.2466/06.08.13.PR0.107.4.120-126

Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M.-L., & Pulkkinen, L. (2006). The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. *Psychological Assessment, 18*, 444-451. doi:10.1037/1040-3590.18.4.444

Malakouti, S. K., Fatollahi, P., Mirabzadeh, A., & Zandi, T. (2007). Reliability, validity and factor structure of the GHQ-28 used among elderly Iranians. *International Psychogeriatrics, 19*, 623-634. doi:10.1017/S1041610206004522

Mansor, M., Haque, M., Sheikh, S. A., Choon, L. C., & Zin, A. M. (2016). Reliability and factor analysis of General Health Questionnaire 12 in Malay version among women

with abnormal pap smear results. *Australasian Medical Journal, 9*, 357-364. doi:10.4066/AMJ.2016.2711

Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. *Psychological Reports, 84*, 927-935. doi:10.2466/pr0.1999.84.3.927

Martin, C. R., & Newell, R. J. (2005). The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement. *Journal of Psychosomatic Research, 59*, 193-199. doi:10.1016/j.jpsychores.2005.02.020

Medina-Mora, M. E., Padilla, G. P., Campillo-Serrano, C., Mas, C. C., Ezban, M., Caraveo, J., & Corona, J. (1983). The factor structure of the GHQ: a scaled version for a hospital's general practice service in Mexico. *Psychological Medicine, 13*, 355-361. doi:10.1017/S0033291700050984

Meertens, V. P. (2004). *Depressive symptoms in the general population: A multifactorial social approach* (Professorial thesis, Katholieke Universiteit Nijmegen, The Netherlands). Retrieved from http://hdl.handle.net/2066/19441

Molavi, H. (2002). Validation, factor structure, and reliability of the Farsi version of the General Health Questionnaire-18 on Irani students. *Pakistan Journal of Psychological Research, 17*, 87-98.

Molina, J. D., & Andrade, C. (2002). The factor structure of the GHQ-60 in a community sample: a scaled version for the Spanish population. *Avances en salud mental relacional, 1*:2. Retrieved from http://www.bibliopsiquis.com/asmr/0102/0102lef2.htm

Molina, J. D., Andrade-Rosa, C., González-Parra, S., Blasco-Fontecilla, H., Real, M. A., & Pintor, C. (2006). The factor structure of the General Health Questionnaire (GHQ): a scaled version for general practice in Spain. *European Psychiatry, 21*, 478-486. doi:10.1016/j.eurpsy.2006.03.002

Molina, J. G., Rodrigo, M. F., Losilla, J.-M., & Vives, J. (2014). Wording effects and the factor structure of the 12-Item General Health Questionnaire (GHQ-12). *Psychological Assessment, 26*, 1031-1037. doi:10.1037/a0036472

Montazeri, A., Harirchi, A. M., Shariati, M., Garmaroudi, G., Ebadi, M., & Fateh, A. (2003). The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. *Health and Quality of Life Outcome, 1*:66. doi:10.1186/1477-7525-1-66

Monteiro, A. P. T. (2011). Assessment of the factor structure and reliability of the Portuguese version of the General Health Questionnaire-28 among adults. *Journal of Mental Health, 20*, 15-20. doi:10.3109/09638237.2010.492414

Munyombwe, T., West, R. M., & Hill, K. (2015). Testing measurement invariance of the GHQ-28 in stroke patients. *Quality of Life Research, 24*, 1823-1827. doi:10.1007/s11136-015-0924-8

Nagyova, I., Krol, B., Szilasiova, A., Stewart, R. E., van Dijk, J. P., & van den Heuvel, W. J. A. (2000). General Health Questionnaire-28: psychometric evaluation of the Slovak version. *Studia Psychologica, 42*, 351-361.

Najarkolaei, F. R., Raiisi, F., Rahnama, P., Fesharaki, M. G., Zamani, O., Jafari, M. R., & Montazeri, A. (2014). Factor structure of the Iranian version of 12-Item General Health Questionnaire. *Iranian Red Crescent Medical Journal, 16*:e11794. doi:10.5812/ircmj.11794

Netz, Y., Dunsky, A., Zach, S., Goldsmith, R., Shimony, T., Goldbourt, U., & Zeev, A. (2012). Psychological functioning and adherence to the recommended dose of physical activity in later life: results from a national health survey. *International Psychogeriatrics, 24*, 2027-2036. doi:10.1017/S1041610212001299

O'Connor, D. W., & Parslow, R. A. (2010). Mental health scales and psychiatric diagnoses: Responses to GHQ-12, K-10 and CIDI across the lifespan. *Journal of Affective Disorders, 121*, 263-267. doi: 10.1016/j.jad.2009.06.038

Ohta, Y., Kawasaki, N., Araki, K., Mine, M., & Honda, S. (1995). The factor structure of the General Health Questionnaire (GHQ-30) in Japanese middle-aged and elderly residents. *International Journal of Social Psychiatry, 41*, 268-275. doi:10.1177/002076409504100404

Ohta, Y., Kawasaki, N., Araki, K., Kawasaki, N., Nakane, Y., Honda, S., & Mine, M. (1998). Psychological trauma and longitudinal course of psychiatric problems among evacuees of a volcanic eruption. *Psychiatry and Clinical Neurosciences, 52* (Suppl.), S41-S48. doi:10.1046/j.1440-1819.1998.0520s1S41.x

Ohta, Y., Araki, K., Kawasaki, N., Nakane, Y., Honda, S., & Mine, M. (2003). Psychological distress among evacuees of a volcanic eruption in Japan: A follow-up study. *Psychiatry and Clinical Neurosciences, 57*, 105-111.

Ozaki, K., Nagata, K., Hayashida, M., Ohta, Y., Ishii, N., & Takemoto, T.-I. (2002). 30-Item General Health Questionnaire scores in male and female university freshmen. *Acta medica Nagasakiensia, 47*, 15-22.

Padrón, A., Galán, I., Durbán, M., Gandarillas, A., & Rodríguez-Artalejo, F. (2012). Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Quality of Life Research, 21*, 1291-1298. doi: 10.1007/s11136-011-0038-x

Pariente, P. D., Challita, H., Mesbah, M., & Guelfi, J. D. (1992). The GHQ-28 questionnaire in French: A validation survey in a panel of 158 general psychiatric patients. *European Psychiatry, 7*, 15-20.

Parkes, K. R. (1982). Field dependence and the factor structure of the General Health

    Questionnaire in normal subjects. *British Journal of Psychiatry, 140*, 392-400.

    doi:10.1192/bjp.140.4.392

Petkovska, M.S., Bojadziev, M. I., & Stefanovska, V. V. (2015). Reliability, validity and

    factor structure of the 12-item General Health Questionnaire among general population.

    Macedonian Journal of Medical Sciences, 3, 478-483. doi:10.3889/oamjms.2015.075

Picardi, A., Abeni, D., & Pasquini, P. (2001). Assessing psychological distress in patients

    with skin diseases: reliability, validity and factor structure of the GHQ-12. *Journal of

    the European Academy of Dermatology and Venereology, 15*, 410-417.

    doi:10.1046/j.1468-3083.2001.00336.x

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure

    of the 12-item General Health Questionnaire among young males in Italy. *Acta

    Psychiatrica Scandinavica, 90*, 432-437. doi:10.1111/j.1600-0447.1994.tb01620.x

Prady, S. L., Miles, J. N., Pickett, K. E., Fairley, L., Bloor, K., Gilbody, S., ... & Wright, J.

    (2013). The psychometric properties of the subscales of the GHQ-28 in a multi-ethnic

    maternal sample: results from the Born in Bradford cohort. *BMC Psychiatry, 13*:55.

    doi:10.1186/1471-244X-13-55

Rajabi, G., & Sheykhshabani, S. H. (2009). Factor structure of the 12-Item General Health

    Questionnaire. *Journal of Education & Psychology, 3*, 81-94.

Rivas-Diez, R., & Sánchez-López, M. P. (2014). Propiedades psicoétricas del Cuestionario de

    Salud General (GHQ-12) en población femenina chilena [Psychometric properties of the

    General Health Questionnaire (GHQ-12) in a Chilean female population]. *Revista

    Argentina de Clínica Psicologica, 23*, 251-260.

Rocha, K. B., Pérez, K., Rodríguez-Sanz, M., Borrell, C., & Obiols, J. E. (2011). Propiedades

     psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en

     población general Española [Psychometric properties und norm values of the General

     Health Questionnaire (GHQ-12) in a Spanish community population]. *International*

     Journal of Clinical and Health Psychology, 11, 125-139.

Romppel, M., Braehler, E., Roth, M., & Glaesmer, H. (2013). What is the General Health

     Questionnaire-12 assessing? Dimensionality and psychometric properties of the

     General HealthQuestionnaire-12 in a large scale German population sample.

     Comprehensive Psychiatry, 54, 406-413. doi:10.1016/j.comppsych.2012.10.010

Salama-Younes, M., Montazeri, A., Ismaïl, A., & Roncin, C. (2009). Factor structure and

     internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the

     Subjective Vitality Scale (VS), and the relationship between them: a study from

     France. *Health and Quality of Life Outcomes, 7*:22. doi:10.1186/1477-7525-7-22

Sánchez-Lopéz, M. P., & Dresch, V. (2008). The 12-Item General Health Questionnaire

     (GHQ-12): Reliability, external validity and factor structure in the Spanish population.

     Psicothema, 20, 839-843.

Shek, D. T. L. (1987). Reliability and factorial structure of the Chinese version of the General

     Health Questionnaire. *Journal of Clinical Psychology, 43*, 683-691. doi:10.1002/1097-

     4679(198711)43:6<683::AID-JCLP2270430606>3.0.CO;2-B

Shek, D. T. L. (1993). Factor structure of the Chinese version of the General Health

     Questionnaire (GHQ-30): A confirmatory factor analysis. *Journal of Clinical*

     *Psychology, 4*9, 678-684. doi:10.1002/1097-4679(199309)49:5<678::AID-

     JCLP2270490510>3.0.CO;2-H

Shek, D. T. L., & Tsang, S. K. M. (1995). Reliability and factor structure of the Chinese

GHQ-30 for parents preschool mentally handicapped children. *Journal of Clinical*

Psychology, 51, 227-234. doi:10.1002/1097-4679(199503)51:2<227::AID-

JCLP2270510212>3.0.CO;2-P

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the

GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychological*

Assessment, 17, 231-236. doi:10.1037/1040-3590.17.2.231

Shigemi, J., Mino, Y., & Tsuda, T. (2000). Stability of factor structure and correlation with

perceived job stress in General Health Questionnaire: a three-wave survey over one

year in Japanese workers. *Journal of Occupational Health, 42*, 281-291.

doi:10.1539/joh.42.284

Siegert, R. J., McCormick, I. A., Taylor, A. J. W., & Walkey, F. H. (1987). An examination

of reported factor structures of the General Health Questionnaire and the identification

of a stable replicable structure. *Australian Journal of Psychology, 39*, 89-100.

doi:10.1080/00049538708259038

Siegert, R. J., & Chung, R. C.-Y. (1995). Dimensions of distress: A cross-cultural factor

replication. *Journal of Cross-Cultural Psychology, 26*, 169-175.

doi:10.1177/0022022195262004

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor

structure of the GHQ-12: the interaction between item phrasing, variance and levels of

distress. *Quality of Life Research*, *22*, 145-152. doi:10.1007/s11136-012-0133-7

Stevenson, J., Simpson, J., & Bailey, V. (1989). The factor structure of the GHQ-30 for

mothers with young children. *Journal of Reproductive and Infant Psychology, 7*, 39-

46. doi:10.1080/02646838908403569

Stuart, G. W., Klimidis, S., Minas, I. H., & Tuncer, C. (1993). The factor structure of the

Turkish version of the General Health Questionnaire. *International Journal of Social*

Psychiatry, 39, 274-284. doi:10.1177/002076409303900404

Sun, J., Stewart, D., Yuan, B. J. & Zhang, S. H. (2012). Validation and normalization of the

General Health Questionnaire 30 in parents with primary school children in China.

Comprehensive Psychiatry, 53, 593-599. doi:10.1016/j.comppsych.2011.09.009

Suzuki, H., Kaneita, Y., Osaki, Y., Minowa, M., Kanda, H., Suzuki, K., …, & Ohida, T.

(2011). Clarification of the factor structure of the 12-item General Health

Questionnaire among Japanese adolescents and associated sleep status. *Psychiatry*

Research, 188, 138-146. doi:10.1016/j.psychres.2010.10.025

Sveen, U., Thommessen, B., Bautz-Holter, E., Wyller, T. B., & Laake, K. (2004). Well-being

and instrumental activities of daily living after stroke. *Clinical Rehabilitation, 18*, 267-

274. doi:10.1191/0269215504cr719oa

Takeuchi, M., & Kitamura, T. (1991). The factor structure of the General Health

Questionnaire in a Japanese high school and university student sample. *International*

Journal of Social Psychiatry, 37, 99-106. doi:10.1177/002076409103700204

Talwar, P., & Rahman, M. F. A. (2015). Factor structure of General Health Questionnaire and

Assessment: A cross-sectional study among university students in Malaysia. *ASEAN*

Journal of Psychiatry, 16, 232-240.

Thabet, A. A. M. (2005). Validity of the Arabic version of the General Health Questionnaire

in the Gaza strip. *Palestinian Medical Journal, 1*, 33-36.

Tomás, J. M., Gutiérrez, M., & Sancho, P. (2015). Factorial validity of the General Health

Questionnaire 12 in an Angolan sample. *European Journal of Psychological*

Assessment, 33, 116-122. doi:10.1027/1015-5759/a000278.

Toyabe, S.-I., Shioiri, T., Kuwabara, H., Endoh, T., Tanabe, N., Someya, T., & Akazawa, K. (2006). Impaired psychological recovery in the elderly after the Niigata-Chuetsu earthquake in Japan: a population-based study. *BMC Public Health, 6*:230 doi:10.1186/1471-2458-6-230

Urzúa, A., Caqueo-Urízar, A., Bargsted, M., & Irarrázaval, M. (2015). ¿Afecta la forma de puntuación la estructurafactorial del GHQ-12? Estudio exploratorioen estudiantes iberoamericanos [Does the GHQ-12 scoring system affect its factor structure? An exploratory study of Ibero American students]. *Cadernos de Saúde Pública, 31*, 1305-1312. doi:10.1590/0102-311X00122913

Vallejo, M. A., Jordán, C. M., Díaz, M. I., Comeche, M. I., & Ortega, J. (2007). Psychological assessment via the Internet: A reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *Journal of Medical Internet Research, 9*:e1. doi:10.2196/jmir.9.1.e2

Vanheule, S., & Bogaerts, S. (2005). The factorial structure of the GHQ-12. *Stress and Health, 21*, 217-222. doi:10.1002/smi.1058

Vázquez-Barquero, J. L., Williams, P., Díez-Manrique, J. F., Lequerica, J., & Arenal, A. (1988). The factor structure of the GHQ-60 in a community sample. *Psychological Medicine, 18*, 211-218. doi:10.1017/S0033291700002038

Villa, I. C., Arboleda, C. Z., & Roldán, L. F. R. (2013). Propiedades psicométricas del Cuestionariode Salud General de Goldberg GHQ-12 en una institución hospitalaria de la ciudad de Medellín [Psychometric properties of the Goldberg's General Health Questionnaire GHQ-12 applied at a hospital facility in the city of Medellin]. *Avances en Psicología Latinoamericana, 31*, 532-545.

Wang, L., & Lin, W. (2011). Wording effects and the dimensionality of the General Health

    Questionnaire (GHQ-12). *Personality and Individual Differences, 50,* 1056-1061.

    doi:10.1016/j.paid.2011.01.024

Weyerer, S., Elton, M., Diallina, M., & Fichter, M. M. (1986). The principal component

    structure of the General Health Questionnaire among Greek and Turkish adolescents.

    European Archives of Psychiatry and Neurological Sciences, 236, 75-82.

    doi:10.1007/BF00454015

Winzer, R. , Lindblad, F., Sorjonen, K., & Lindberg, L. (2014). Positive versus negative

    mental health in emerging adulthood: a national cross-sectional survey. *BMC Public

    Health, 14*:1238. doi: 10.1186/1471-2458-14-1238

Wong, K. C. K., & O`Driscoll, M. P. (2016). Psychometric properties of the General Health

    Questionnaire-12 in a sample of Hong Kong employees. *Psychology, Health &

    Medicine.* Advance online publication. doi: 10.1080/13548506.2016.1140901

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of

    wording effects. *Personality and Individual Differences, 46,* 197-201.

    doi:10.1016/j.paid.2008.09.027

Zelča, I., Žermēna, V. Čuprika, A., Amantova, I., Galeja, Z., Ozoliņa, L., Vaisjūns, O., &

    Ozola, M. (2013). Psychometric indicators of General Health Questionnaire in Latvia.

    LASE Journal of Sport Science, 4, 205-214.

Zulkefly, N. S., & Baharudin, R. (2010). Using the 12-item General Health Questionnaire

    (GHQ-12) to assess the psychological health of Malaysian college students. *Global

    Journal of Health Science, 2,* 73-80.

Studies Included in Meta-Analysis I

Abeysena, C., Jayawardana, P., & Peiris, U. (2012). Factor structure and reliability of the 12-
item Sinhala version of General Health Questionnaire. *International Journal of
Collaborative Research on Internal Medicine & Public Health, 4,* 1606-1613.

Borges, L. O., & Argolo, J. C. T. (2002). Adaptação e validação de uma escala de bem-estar
psicológico para uso em estudos ocupacionais [Adaptation and validation of a
psychological well-being scale for use in occupational studies]. *Avaliação Psicológica, 1,*
17-27.

Bratås, O., Grønning, K., & Forbord, T. (2013). Psychometric properties of the Hospital
Anxiety and Depression Scale and the General Health Questionnaire-20 in COPD
inpatients. *Scandinavian Journal of Caring Sciences, 28,* 413-420. doi:10.1111/scs.12054

Burrone, M. S., Abeldaño, A., Lucchese, M., Susser, L., Enders, J. E., Alvarado, R., ... &
Fernández, A. R. (2015). Evaluación psicométrica y estudio de fiabilidad del cuestionario
general de salud (GHQ-12) en consultantes adultos del primer nivel de atención en
Córdoba - argentina [Psychometric properties and reliability of the General Health
Questionnaire (GHQ-12) for adult patients in primary care centers in Cordoba,
Argentina]. *Revista de la Facultad de Ciencias Medicas*, *73,* 236-242.

Carné, J. A. (2004). Anàlisi factorial confirmatòria amb indicadors categòrics: Estudi de la
dimensionalitat del GHQ-12 [A confirmatory factor analysis with categorical indicators:
A study of the dimensionality of the GHQ-12]. (Master's thesis, Universitat Politècnica
Catalunya, Spain). Retrieved from http://hdl.handle.net/2099.1/13755

Chan, D. W., & Chan, T. S. (1983). Reliability, validity and the structure of the General
Health Questionnaire in a Chinese context. *Psychological Medicine, 13,* 363-371.
doi:10.1017/S0033291700050996

Cuéllar-Flores, I., Pilar Sánchez-López, M., María Limiñana-Gras, M., & Colodro-Conde, L.

    (2014). The GHQ-12 for the assessment of psychological distress of family caregivers.

    Behavioral Medicine, 40, 65-70. doi:10.1080/08964289.2013.847815

D'Arcy, C., & Siddique, C. M. (1984). Psychological distress among Canadian adolescents.

    Psychological Medicine, 14, 615-628. doi:10.1017/S0033291700015221

Dale, B., Sævareid, H. I., & Söderhamn, O. (2009). Testing and using Goldberg's General

    Health Questionnaire: Mental health in relation to home nursing, home help, and family

    care among older, care-dependent individuals. *International Journal of Mental Health*

    Nursing, 18, 133-143. doi:10.1111/j.1447-0349.2009.00593.x

Damásio, B. F., Machado, W. L., & Silva, J. P. (2011). Estrutura fatorial do questionário de

    saúde geral (QSG-12) em uma amostra de professores escolares [Factor structure of the

    General Health Questionnaire (GHQ-12) in a sample of school teachers]. *Avaliação*

    Psicológica, 10, 99-105.

Frydecka, D., Małyszczak, K., Chachaj, A., & Kiejna, A. (2010). Struktura czynnikowa

    Kwestionariusza Ogólnego Zdrowia (GHQ-30) [Factorial structure of the General Health

    Questionnaire (GHQ-30)]. *Psychiatria Polska, 44*, 341-359.

Garmendia, M. L. (2007). Análisis factorial: una aplicación en el cuestionario de salud

    general de Goldberg, versión de 12 preguntas [Factor analysis: an application to

    Goldberg's General Health Questionnaire, 12-item version]. *Revista Chilena de Salud*

    Pública, 11, 57-65. doi:10.5354/0719-5281.2007.3095

Glozah, F. N., & Pevalin, D. J. (2015). Factor structure and psychometric properties of the

    General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *Journal of Child*

    & Adolescent Mental Health, 27, 53-57. doi:10.2989/17280583.2015.1007867

Gouveia, V. V., Chaves, S. S., Oliveira, I. C. P., Dias, M. R., Gouveia, R. S. V., & Andrade,

P. R. (2003). A utilização do QSG-12 na população geral: Estudo de sua validade de

construto [The use of the GHQ-12 in the general population: A study on its construct

validity]. *Psicologia: Teoria e Pesquisa. 19*, 241-248. doi:10.1590/S0102-

37722003000300006

Gouveia, V. V., de Lima, T. J. S., Gouveia, R. S. V., Freires, L. A., & Barbosa, L. H. G. M.

(2012). Questionário de Saúde Geral (QSG-12): o efeito de itens negativos em sua

estrutura factorial [General Health Questionnaire (GHQ-12): The effect of negative items

on its factorial structure]. *Cadernos de Saúde Pública, 28*, 375-384. doi:10.1590/S0102-

311X2012000200016

Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social

Psychiatry and Psychiatric Epidemiology, 26*, 132-138. doi:10.1007/BF00782952

Gureje, O. (1991). Reliability and the factor structure of the Yoruba version of the 12-item

General Health Questionnaire. *Acta Psychiatrica Scandinavica, 84*, 125-129.

doi:10.1111/j.1600-0447.1991.tb03115.x

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire

(GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in

Mental Health, 4*(10). doi:10.1186/1745-0179-4-10

Hu, Y., Stewart-Brown, S., Twigg, L., & Weich, S. (2007). Can the 12-item General Health

Questionnaire be used to measure positive mental health? *Psychological Medicine, 37*,

1005-1013. doi:10.1017/S0033291707009993

Hubbert, F. A., Walters, D. E., Day, N. E., & Elliott, B. J. (1989). The factor structure of the

General Health Questionnaire (GHQ-30). A reliability study on 6317 community

residents. *British Journal of Psychiatry, 155*, 178-185. doi:10.1192/bjp.155.2.178

Ibrahim, A. (2008). *Evaluating training effectiveness in the Malaysian public service*
     (Doctoral thesis, Durham University, United Kingdom). Retrieved from
     http://etheses.dur.ac.uk/2176/

Ip, W. Y., & Martin, C. R. (2006). Factor structure of the Chinese version of the 12-item
     General Health Questionnaire (GHQ-12) in pregnancy. *Journal of Reproductive and
     Infant Psychology, 24*, 87-98. doi:10.1080/02646830600643882

Iwata, N., Okuyama, Y., Kawakami, Y., & Saito, K. (1988). The twelve-item General Health
     Questionnaire among Japanese workers. *Environmental Science, 11*, 1-10.

Iwata, N., Uno, B., & Suzuki, T. (1994). Psychometric properties of the 30-item version
     General Health Questionnaire in Japanese. *Psychiatry and Clinical Neurosciences, 48*,
     547-556. doi:10.1111/j.1440-1819.1994.tb03013.x

Jacob, K. S., & Bhugra, D., & Mann, A. H. (1997). General health questionnaire - 12:
     Psychometric properties and factor structure among Indian women living in the United
     Kingdom. *Indian Journal of Psychiatry, 39*, 196-199.

Khan, A., Shah, I. M., Khan, F., & Suhail, S. (2013). Reliability and validity assessment of 12
     items General Health Questionnaire (GHQ: 12) among Pakistani university teachers.
     World Applied Sciences Journal, 24, 603-608.

Kihç, C., Rezaki, M., Rezaki, B., Kaplan, I., Özgen, G., Sagduyu, A., & Özturk, M. O.
     (1997). General Health Questionnaire (GHQ12 & GHQ28): Psychometric properties and
     factor structure of the scales in a Turkish primary care sample. *Social Psychiatry and
     Psychiatric Epidemiology, 32*, 327-331. doi:10.1007/BF00805437

Kuruvilla, A., Pothen, M., Philip, K., Braganza, D., Joseph, A., & Jacob, K. S. (1999). The
     validation of the Tamil version of the 12 item General Health Questionnaire. *Indian
     Journal of Psychiatry, 41*, 217-221.

López-Castedo, A. & Fernández, L. (2005). Psychometric properties of the Spanish version of the 12-item General Health Questionnaire in adolescents. *Perceptual and Motor Skills, 100,* 676-680. doi:10.2466/pms.100.3.676-680

Oliveira, G. F. (2008). *Trabalho e bem-estar subjetivo: Compreendendo a situação laboral dos medicos* [Work and subjective well-being: Understanding the work conditions of doctors] (Doctoral thesis, Federal University of Paraíba, Brazil). Retrieved from http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?co_obra=146584

Park, J.-I., Kim, Y. J., & Cho, M. J. (2012). Factor structure of the 12-Item General Health Questionnaire in the Korean general adult population. *Journal of the Korean Neuropsychiatric Association, 51,* 178-184. doi:10.4306/jknpa.2012.51.4.178

Penninkilampi-Kerola, V., Miettunen, J. & Ebeling, H. (2007). 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. *Scandinavian Journal of Psychology, 47,* 431-440. doi:10.1111/j.1467-9450.2006.00551.x

Piccinelli, M., & Politi, P. (1993). Struttura fattoriale della versione a 12 domande del General Health Questionnaire in un campione di giovani maschi adulti [Factor structure of the 12-items version of the General Health Questionnaire in a sample of young adult males]. *Epidemiologia e Psychiatria Sociale, 2,* 173-181. doi:10.1017/S1121189X00006990

Sarková, M., Nagyová, I., Katreniaková, Z., Madrasová Gecková, A., Orosová, O., Middel, B., van Dijk, J. P., & van den Heuvel, W. (2006). Psychometric evaluation of the General Health Questionnaire – 12 and Rosenberg Self-Esteem Scale in Hungarian and Slovak early adolescents. *Studia Psychologica, 48,* 69-79.

Solís-Cámara, P., Lara, R. M. M., Jiménez, B. M., & Rodríguez, P. J. (2016). Estructura factorial del cuestionario de salud general GHQ-12 en población general de México

[Factor structure of the GHQ-12 general health questionnaire in Mexican population]. *Salud & Sociedad*, *7*, 62-76.

Viniegras, C. R. V. G. (1999). Manual para la utilización del Cuestionario de Salud General de Goldberg. Adaptación cubana [Manual for using the Goldberg's General Health Questionnaire. Cuban adapation]. *Revista Cubana de Medicina General Integral, 15*, 88-97.

Wang, W. C. (2005). *A comparison of alternative estimation methods in confirmatory factor analyses of the General Health Questionnaire across four groups of Australian immigrants* (Master's thesis, Swinburne University of Technology, Australia). Retrieved from http://hdl.handle.net/1959.3/25921

Yasuda, N., Mino, Y., Koda, S., & Ohara, H. (2002). The differential influence of distinct clusters of psychiatric symptoms, as assessed by the General Health Questionnaire, on cause of death in older persons living in a rural community of Japan. *Journal of the American Geriatrics Society, 50*, 313-320. doi:10.1046/j.1532-5415.2002.50064.x

Zangmo, T. (2013). Factorial validity and reliability of 12 items General Health Questionnaire in a Bhutanese population. *Journal of Bhutan Studies, 29*, 109-134.

Characteristics of Studies Included in Meta-Analysis I

Study	Year	Country	*N*	Sex	Age	Items	Statistic	Factors	Method	Rotation	Missing
Abeysena et al. (2012)	2012	Sri Lanka	385		32	12	Loading	2	PAF	Oblique	50
Borges & Argolo (2002)	2002	Brazil	446	54	35	12	Loading	2	PAF	Oblique	50
Bratås et al. (2013)	2013	Norway	161			20	Loadings	3	PAF	Oblique	0
Burrone et al. (2015)	2015	Argentina	854	83	35	12	Loadings	2	PCA	Orthogonal	50
Carné (2004)	2004	Spain	380	100		12	Correlation				
Cuéllar-Flores et al. (2014)	2014	Spain	172	85	57	12	Loading	2	PCA	Orthogonal	50
Dale et al. (2009)	2009	Norway	234	70	85	30	Loadings	8	PCA	Orthogonal	0
Damásio et al. (2001)	2011	Brazil	203	66	37	12	Loading	2	ML	Oblique	0
D'Arcy (1984)	1984	Canada	1,038	51		30	Loading	4	?	Orthogonal	0
Frydeck et al. (2010)	2010	Poland	623	79		30	Loadings	3	PCA	Orthogonal	31
Garmendia (2007)	2007	Chile	306	20	30	12	Correlation				
Glozah & Pevalin (2015)	2015	Ghana	770	35	17	12	Loading	2	PAF	Oblique	50
Gouveia et al. (2003)	2003	Brazil	306		34	12	Loading	3	PAF	Oblique	0
Gouveia et al. (2012)	2012	Brazil	1,180	50	32	12	Correlation				
Graetz (1991)	1991	Australia	6,151	50		12	Loading	3	ML	Oblique	0

Study	Year	Country	*N*	Sex	Age	Items	Statistic	Factors	Method	Rotation	Missing
Gureje (1991)	1991	Nigeria	787	51		12	Loading	2	PCA	Orthogonal	0
Hankins (2008)	2008	UK	3,705			12	Correlation				
Hu et al. (2007)	2007	UK	8,978	53	42	12	Loading	2	PCA	Orthogonal	0
		UK	6,451	53	43	12	Loading	2	PCA	Orthogonal	0
Hubbert et al. (1989)	1989	UK	6,000			30	Loadings	5	?	Orthogonal	0
Ibrahim (2008)	2008	Malaysia	757	56		12	Correlation				
Ip & Martin (2006)	2006	Hongkong	292	100	30	12	Loading	2	PCA	Oblique	0
Iwata et al. (1988)	1988	Japan	1,927	0		12	Loading	2	PCA	Orthogonal	0
		Japan	143	100		12	Loading	2	PCA	Orthogonal	0
Iwata et al. (1994)	1994	Japan	2,986	18		30	Loadings	6	PCA	Orthogonal	0
Jacob et al. (1997)	1997	UK	125	100	42	12	Loading	2	PCA	Orthogonal	0
Khan et al. (2014)	2014	Pakistan	400			12	Loading	3	PCA	Orthogonal	0
Kihç et al. (1997)	1997	Turkey	1,307	66	34	12	Loading	2	PCA	Orthogonal	0
Kuruvilla et al. (1999)	1999	India	327	66	35	12	Loading	3	PCA	Orthogonal	0
López-Castedo & Fernández (2005)	2005	Spain	1,930	55	16	12	Loading	2	PCA	Orthogonal	50
Oliveira (2008)	2008	Brazil	246			12	Loading	2	PCA	Oblique	0

Study	Year	Country	N	Sex	Age	Items	Statistic	Factors	Method	Rotation	Missing
Park et al. (2012)	2012	South Korea	6,510	60		12	Loading	2	?	Oblique	50
Penninkilampi-Kerola et al. (2007)	2007	Finnland	2,294	56		20	Loading	4	PAF	Oblique	0
Piccinelli & Politi (1993)	1993	Italy	320			12	Loading	2	PCA	Orthogonal	0
Sarková et al. (2006)	2006	Slovakia	519	49	12	12	Loading	2	PCA	Orthogonal	0
		Hungary	431	47	12	12	Loading	2	PCA	Orthogonal	0
Solís-Cámara et al. (2016)	2016	Mexico	1,093	58	31	12	Correlation				
Viniegras (1999)	1999	Cuba	237	41	39	12	Loading	2	PCA	Orthogonal	0
Wang (2005)	2005	Hongkong	201	41	33	12	Correlation				
		China	213	40	36	12	Correlation				
		Yugoslavia	259	39	38	12	Correlation				
		UK	428	28	37	12	Correlation				
Yasuda et al. (2002)	2002	Japan	980	61	72	30	Loadings	5	PCA	Orthogonal	0
Zangmo (2013)	2013	Bhutan	6,861			12	Correlation				

Note. Year = Publication year; Country = Country of origin; Sex = Percentage of female participants; Age = Mean age (in years); Item = GHQ version as number of items; Type = Coded statistic as factor loading pattern or correlation matrix; Factors = Number of extracted factors; Method = Factor extraction method as principal components analysis (PCA), principal axis factoring (PAF), maximum likelihood analysis (ML), or unknown (?); Rotation = Factor rotation method as orthogonal or oblique; Missing = Percentage of missing factor loadings.

Coding Guide for Meta-Analysis I

Variable	Description	Value	Example
STUDY	Study ID: last name of first author + publication year	Open text	carl2007
	In case of multiple samples within on study: append a single letter		carl2007a
PYR	Publication year	Value range: [1900, 2015]	2007
SYR	Survey year	Value range: [1900, 2015]	2007
CNTRY	Country of origin of participants	Open text	Germany
LANG	Language of questionnaire	Open text	German
PUB	Publication type	1 = Peer-reviewed journal	1
		2 = Book	
		3 = Thesis (Master / PhD)	
		4 = Presentation / Proceedings	
		5 = other	
N	Sample size	Value range: [2, ∞]	100
SEX	Percentage of women in sample (%)	Value range: [0, 1]	.40
AGE	Mean age (in years) of participant	Value range: [13, ∞]	16.00
CAT	Number of response categories	Open text	4
SCORE	How were items scored?	1 = Likert scale	1
		2 = Dichotomous	
		3= other	
SCORE2	How were items scored? other	Open text	
FAC	Number of extracted factors	Value range: [1, 12]	1
MET	Factoranalytic method	1 = Principal component	1
		2 = Principal axis	
		3 = Maximum likelihood	
		4 = other	
MET2	Factoranalytic method: other	Open text	
ROT	Rotation method	1 = varimax	1
		2 = oblimin	
		3 = promax	
		4 = other	
ROT2	Rotation method: other	Open text	
F1.1	Factor 1 - item 1: Able to concentrate	Value range: [-1, 1]	.50
F1.2	Factor 1 – item 2: Lost much sleep over worry	Value range: [-1, 1]	.50
F1.3	Factor 1 – item 3: Playing a useful part	Value range: [-1, 1]	.50
F1.4	Factor 1 – item 4: Capable of making decisions	Value range: [-1, 1]	.50
F1.5	Factor 1 – item 5: Felt constantly under stress	Value range: [-1, 1]	.50
F1.6	Factor 1 – item 6: Could not overcome difficulties	Value range: [-1, 1]	.50
F1.7	Factor 1 – item 7: Enjoy day-to-day (normal) activities	Value range: [-1, 1]	.50
F1.8	Factor 1 – item 8: Able to face up to problems	Value range: [-1, 1]	.50

Variable	Description	Value	Example
F1.9	Factor 1 – item 9: Feeling unhappy and depressed	Value range: [-1, 1]	.50
F1.10	Factor 1 – item 10: Losing confidence	Value range: [-1, 1]	.50
F1.11	Factor 1 – item 11: Thinking of self as worthless	Value range: [-1, 1]	.50
F1.12	Factor 1 – item 12: Feeling reasonably happy	Value range: [-1, 1]	.50
F2.1	Factor 2 - item 1: Able to concentrate	Value range: [-1, 1]	.50
F2.2	Factor 2 – item 2: Lost much sleep over worry	Value range: [-1, 1]	.50
F2.3	Factor 2 – item 3: Playing a useful part	Value range: [-1, 1]	.50
F2.4	Factor 2 – item 4: Capable of making decisions	Value range: [-1, 1]	.50
F2.5	Factor 2 – item 5: Felt constantly under stress	Value range: [-1, 1]	.50
F2.6	Factor 2 – item 6: Could not overcome difficulties	Value range: [-1, 1]	.50
F2.7	Factor 2 – item 7: Enjoy day-to-day (normal) activities	Value range: [-1, 1]	.50
F2.8	Factor 2 – item 8: Able to face up to problems	Value range: [-1, 1]	.50
F2.9	Factor 2 – item 9: Feeling unhappy and depressed	Value range: [-1, 1]	.50
F2.10	Factor 2 – item 10: Losing confidence	Value range: [-1, 1]	.50
F2.11	Factor 2 – item 11: Thinking of self as worthless	Value range: [-1, 1]	.50
F2.12	Factor 2 – item 12: Feeling reasonably happy	Value range: [-1, 1]	.50
F3.1	Factor 3 - item 1: Able to concentrate	Value range: [-1, 1]	.50
F3.2	Factor 3 – item 2: Lost much sleep over worry	Value range: [-1, 1]	.50
F3.3	Factor 3 – item 3: Playing a useful part	Value range: [-1, 1]	.50
F3.4	Factor 3 – item 4: Capable of making decisions	Value range: [-1, 1]	.50
F3.5	Factor 3 – item 5: Felt constantly under stress	Value range: [-1, 1]	.50
F3.6	Factor 3 – item 6: Could not overcome difficulties	Value range: [-1, 1]	.50
F3.7	Factor 3 – item 7: Enjoy day-to-day (normal) activities	Value range: [-1, 1]	.50
F3.8	Factor 3 – item 8: Able to face up to problems	Value range: [-1, 1]	.50
F3.9	Factor 3 – item 9: Feeling unhappy and depressed	Value range: [-1, 1]	.50
F3.10	Factor 3 – item 10: Losing confidence	Value range: [-1, 1]	.50
F3.11	Factor 3 – item 11: Thinking of self as worthless	Value range: [-1, 1]	.50
F3.12	Factor 3 – item 12: Feeling reasonably happy	Value range: [-1, 1]	.50
R1.2	Correlation between Factor 1 and 3	Value range: [-1, 1]	.50
R1.3	Correlation between Factor 1 and 3	Value range: [-1, 1]	.50
R2.3	Correlation between Factor 2 and 3	Value range: [-1, 1]	.50
NOTE	General comments	Open text	

Congruence of Bifactor Structures Across Subgroups in Meta-Analysis I

	General factor	Specific factor 1	Specific factor 2	All factors
Coded statistic:				
Correlation vs. full factor loadings	1.00	.90	1.00	.99
Correlation vs. imputed factor loadings	.99	.93	.98	.98
Full vs. imputed factor loading matrix	.99	.87	.98	.98
GHQ version:				
GHQ-12 vs. GHQ-20 / GHQ-30	1.00	.79	.96	.97
Language version:				
English vs. Spanish	1.00	.98	.99	.98
English vs. Portuguese	1.00	.98	.98	.99
English vs. Japanese	.99	.96	.97	.98
Spanish vs. Portuguese	1.00	.95	.99	.97
Spanish vs. Japanese	.99	.93	.99	.98
Portuguese vs. Japanese	.99	.89	.96	.96
Median	1.00	.94	.98	.98

Note. Congruence indices for individual factors (Tucker, 1951) and full factor loading matrices (Gebhardt, 1968).

Datasets Included in Meta-Analysis II

Access Research Knowledge (ARK) Northern Ireland (2012). *Young Life and Times Survey,*
    2011 [computer file]. Colchester, United Kingdom: UK Data Archive [distributor]. SN:
    7058. doi:10.5255/UKDA-SN-7058-1

Department of Health, Social Services and Public Safety (Northern Ireland). (2013). *Northern*
    Ireland Health Survey, 2010-2011 [computer file]. Colchester, United Kingdom: UK
    Data Archive [distributor]. SN: 7258. doi:10.5255/UKDA-SN-7258-1

Devine, P. (2007). *Northern Ireland Life and Times Survey, 2006* [computer file]. Colchester,
    United Kingdom: UK Data Archive [distributor]. SN: 5695. doi:10.5255/UKDA-SN-
    5695-1

Devine, P., & Schubotz, D. (2005). *Young Life and Times Survey, 2004* [computer file].
    Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5175].
    doi:10.5255/UKDA-SN-5175-1

Devine, P., & Schubotz, D. (2006). *Young Life and Times Survey, 2005* [computer file].
    Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5338.
    doi:10.5255/UKDA-SN-5338-1

Hendry, L. B., Shucksmith, J., McCrae, J., & Love, J. (1990). *Young People's Leisure and*
    Lifestyles in Modern Scotland, 1987 [computer file]. Colchester, United Kingdom: UK
    Data Archive [distributor]. SN: 2713. doi:10.5255/UKDA-SN-2713-1

Joint Health Surveys Unit & University College London (2011). *Scottish Health Survey, 2003*
    [computer file] (2nd edition). Colchester, United Kingdom: UK Data Archive
    [distributor]. SN: 5318. doi:10.5255/UKDA-SN-5318-1

Joint Health Surveys Unit of Social and Community Planning Research & University College
    London (1999). *Scottish Health Survey, 1995* [computer file] (3rd edition). Colchester,

United Kingdom: UK Data Archive [distributor]. SN: 3807. doi:10.5255/UKDA-SN-3807-1

Joint Health Surveys Unit of Social and Community Planning Research & University College London (2001). *Health Survey for England, 1994* [computer file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 3640. doi:10.5255/UKDA-SN-3640-1

Joint Health Surveys Unit of Social and Community Planning Research & University College London (2001). *Scottish Health Survey, 1998* [computer file]. Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4379. doi:10.5255/UKDA-SN-4379-1

Joint Health Surveys Unit of Social and Community Planning Research & University College London (2010). *Health Survey for England, 1995* [computer file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 3796. doi:10.5255/UKDA-SN-3796-1

Joint Health Surveys Unit of Social and Community Planning Research & University College London (2010). *Health Survey for England, 1997* [computer file] (3rd edition). Colchester, United Kingdom: UK Data Archive [distributor], SN: 3979. doi:10.5255/UKDA-SN-3979-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2010). *Health Survey for England, 1998* [computer file] (5th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4150. doi:10.5255/UKDA-SN-4150-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2010). *Health Survey for England, 1999* [computer

file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4365. doi:10.5255/UKDA-SN-4365-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2010). *Health Survey for England, 2001* [computer file] (3rd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4628. doi:10.5255/UKDA-SN-4628-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2010). *Health Survey for England, 2002* [computer file] (2nd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4912. doi:10.5255/UKDA-SN-4912-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health. (2010). *Health Survey for England, 2003* [computer file] (2nd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5098. doi:10.5255/UKDA-SN-5098-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health. (2010). *Health Survey for England, 2004* [computer file] (2nd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5439. doi:10.5255/UKDA-SN-5439-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2011). *Health Survey for England, 2000* [computer file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4487. doi:10.5255/UKDA-SN-4487-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health. (2011). *Health Survey for England, 2005* [computer

file] (3rd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5675. doi:10.5255/UKDA-SN-5675-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health. (2011). *Health Survey for England, 2006* [computer file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5809. doi:10.5255/UKDA-SN-5809-1

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2013). *Health Survey for England, 2008* [computer file] (4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6397. doi:10.5255/UKDA-SN-6397-2

National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2015). *Health Survey for England, 2009* [computer file] (3rd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6732. doi:10.5255/UKDA-SN-6732-2

NatCen Social Research & University College London, Department of Epidemiology and Public Health (2014). *Health Survey for England, 2012* [computer file]. Colchester, United Kingdom: UK Data Archive [distributor]. SN: 7480. doi:10.5255/UKDA-SN-7480-1

NatCen Social Research & Royal Free and University College Medical School, Department of Epidemiology and Public Health. (2015). *Health Survey for England, 2010* [computer file] (3rd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6986. doi:10.5255/UKDA-SN-6986-3

NHS Health Scotland (2004). *Health Education Population Survey, 1996-1999* [computer

file]. Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4949.

doi:10.5255/UKDA-SN-4949-1

NHS Health Scotland (2004). *Health Education Population Survey, 2001* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4950.

doi:10.5255/UKDA-SN-4950-1

NHS Health Scotland (2004). *Health Education Population Survey, 2002* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4951.

doi:10.5255/UKDA-SN-4951-1

NHS Health Scotland (2004). *Health Education Population Survey, 2003* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 4952.

doi:10.5255/UKDA-SN-4952-1

NHS Health Scotland (2005). *Health Education Population Survey, 2004* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5202.

doi:10.5255/UKDA-SN-5202-1

NHS Health Scotland (2006). *Health Education Population Survey, 2005* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5470.

doi:10.5255/UKDA-SN-5470-1

NHS Health Scotland (2007). *Health Education Population Survey, 2006* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 5713.

doi:10.5255/UKDA-SN-5713-1

NHS Health Scotland (2008). *Health Education Population Survey, 2007* [computer file].

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6023.

doi:10.5255/UKDA-SN-6023-1

Northern Ireland Statistics and Research Agency, Central Survey Unit (2002). *Northern*

   Ireland Health and Social Wellbeing Survey, 1997 [computer file]. Colchester, United

   Kingdom: UK Data Archive [distributor]. SN: 4589. doi:10.5255/UKDA-SN-4589-1

Northern Ireland Statistics and Research Agency, Central Survey Unit (2002). *Northern*

   Ireland Health and Social Wellbeing Survey, 2001 [computer file]. Colchester, United

   Kingdom: UK Data Archive [distributor]. SN: 4590. doi:10.5255/UKDA-SN-4590-1

Northern Ireland Statistics and Research Agency, Central Survey Unit (2007). *Northern*

   Ireland Health and Social Wellbeing Survey, 2005-2006 [computer file]. Colchester,

   United Kingdom: UK Data Archive [distributor]. SN: 5710. doi:10.5255/UKDA-SN-

   5710-1

Northern Ireland Statistics and Research Agency, Central Survey Unit. (2011). *Continuous*

   Household Survey, 2009-2010 [computer file] (2nd edition). Colchester, United Kingdom:

   UK Data Archive [distributor]. SN: 6731. doi:10.5255/UKDA-SN-6731-1

Office for National Statistics, Social and Vital Statistics Division et al. (2005). *Mental Health*

   of Children and Young People in Great Britain, 2004 [computer file]. Colchester, United

   Kingdom: UK Data Archive [distributor]. SN: 5269. doi:10.5255/UKDA-SN-5269-1

Office of Population Censuses and Surveys, Social Survey Division (1997). *Health Survey for*

   England, 1991-1992: Combined Data File [computer file] (2nd edition). Colchester,

   United Kingdom: UK Data Archive [distributor]. SN: 3238. doi:10.5255/UKDA-SN-

   3238-1

Office of Population Censuses and Surveys, Social Survey Division (1997). *Health Survey for*

   England, 1993 [computer file] (2nd edition). Colchester, United Kingdom: UK Data

   Archive [distributor]. SN: 3316. doi:10.5255/UKDA-SN-3316-1

Schubotz, D. (2007). *Young Life and Times Survey, 2006* [computer file]. Colchester, United

    Kingdom: UK Data Archive [distributor]. SN: 5674. doi:10.5255/UKDA-SN-5674-1

Schubotz, D. (2008). *Young Life and Times Survey, 2007* [computer file]. Colchester, United

    Kingdom: UK Data Archive [distributor]. SN: 5818. doi:10.5255/UKDA-SN-5818-1

Schubotz, D. (2009). *Young Life and Times Survey, 2008* [computer file]. Colchester, United

    Kingdom: UK Data Archive [distributor]. SN: 6274. doi:10.5255/UKDA-SN-6274-1

Schubotz, D. (2014). *Young Life and Times Survey, 2013* [computer file]. Colchester, United

    Kingdom: UK Data Archive [distributor]. SN: 7541. doi:10.5255/UKDA-SN-7541-1

ScotCen Social Research (2014). *Scottish Health Survey, 2013* [computer file]. Colchester,

    United Kingdom: UK Data Archive [distributor]. SN: 7594. doi:10.5255/UKDA-SN-

    7594-1

ScotCen Social Research & University College London, Department of Epidemiology and

    Public Health (2012). *Scottish Health Survey, 2010* [computer file]. Colchester, United

    Kingdom: UK Data Archive [distributor]. SN: 6987. doi:10.5255/UKDA-SN-6987-1

ScotCen Social Research, University College London, Department of Epidemiology and

    Public Health & University of Glasgow, MRC/CSO Social and Public Health Sciences

    Unit (2014). *Scottish Health Survey, 2011* [computer file] (3rd edition). Colchester,

    United Kingdom: UK Data Archive [distributor]. SN: 7247. doi:10.5255/UKDA-SN-

    7247-3

ScotCen Social Research, University College London, Department of Epidemiology and

    Public Health & University of Glasgow, MRC/CSO Social and Public Health Sciences

    Unit (2014). *Scottish Health Survey, 2012* [computer file] (2nd edition). Colchester,

    United Kingdom: UK Data Archive [distributor]. SN: 7417. doi:10.5255/UKDA-SN-

    7417-2

Scottish Centre for Social Research & University College London, Department of

Epidemiology and Public Health (2011). *Scottish Health Survey, 2009* [computer file]

(4th edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6713.

doi:10.5255/UKDA-SN-6713-2

Scottish Centre for Social Research & University College London, Department of

Epidemiology and Public Health (2013). *Scottish Health Survey, 2008* [computer file]

(2nd edition). Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6383.

doi:10.5255/UKDA-SN-6383-2

University of Essex. Institute for Social and Economic Research (2010). *British Household*

Panel Survey: Waves 1-18, 1991-2009 [computer file] (7th edition). Colchester, United

Kingdom: UK Data Archive [distributor]. SN: 5151. doi:10.5255/UKDA-SN-5151-1

University of Essex. Institute for Social and Economic Research, & NatCen Social Research

(2014). *Understanding Society: Waves 1-4, 2009-2013* [computer file] (6th edition).

Colchester, United Kingdom: UK Data Archive [distributor]. SN: 6614.

doi:10.5255/UKDA-SN-6614-6

Fit Statistics for Different Confirmatory Factor Models of

different language versions of the GHQ-12

	Model	χ^2	df	CFI	SRMR	RMSEA	AIC	BIC
		Spanish language version						
1.	Single factor model	2,190[*]	54	.790	.185	.089	2,082	1,731
2.	Artifactual factor model	378[*]	48	.968	.035	.037	282	-31
Andrich & van Schoubroeck (1989)								
3a.	Correlated factor model	691[*]	53	.937	.061	.049	585	240
3b.	Bifactor model	201[*]	42	.984	.024	.028	117	-157
Graetz (1991)								
4a.	Correlated factor model	563[*]	51	.950	.056	.045	461	128
4b.	Bifactor model	202[*]	43	.984	.024	.027	116	-164
Martin (1999)								
5a.	Correlated factor model	3,248[*]	51	.686	.262	.112	3,146	2,814
5b.	Bifactor model	1,549[*]	42	.852	.149	.085	1,465	1,191
		Portuguese language version						
1.	Single factor model	2,972[*]	54	.893	.126	.079	2,864	2,482
2.	Artifactual factor model	1,965[*]	48	.930	.086	.068	1,869	1,530
Andrich & van Schoubroeck (1989)								
2a.	Correlated factor model	2,269[*]	53	.919	.099	.069	2,163	1,788
2b.	Bifactor model	1,396[*]	42	.950	.063	.061	1,312	1,015
Graetz (1991)								
3a.	Correlated factor model	1,891[*]	51	.932	.085	.064	1,789	1,428
3b.	Bifactor model	1,357[*]	43	.952	.062	.059	1,271	967
Martin (1999)								
4a.	Correlated factor model	2,708[*]	51	.902	.115	.077	2,606	2,245
4b.	Bifactor model	1,694[*]	42	.939	.079	.067	1,610	1,313
		Japanese language version						
1.	Single factor model	2,073[*]	54	.850	.138	.079	1,965	1,603
2.	Artifactual model	475[*]	48	.968	.033	.038	379	57
Andrich & van Schoubroeck (1989)								
3a.	Correlated factor model	809[*]	53	.944	.056	.049	703	347
3b.	Bifactor model	200[*]	42	.988	.018	.025	116	-165

Model	χ^2	df	CFI	SRMR	RMSEA	AIC	BIC
Graetz (1991)							
4a. Correlated factor model	689[*]	51	.953	.052	.046	587	245
4b. Bifactor model	239[*]	43	.985	.020	.028	153	-135
Martin (1999)							
5a. Correlated factor model	1,813[*]	51	.869	.121	.076	1,711	1,369
5b. Bifactor model	Model did not converge						

Note. CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

* $p < .05$