Esta es la versión de autor del artículo publicado en:
This is an author produced version of a paper published in:

Psychological Assessment 25.1 (2013): 279-287

DOI:  http://dx.doi.org/10.1037/a0030645

Copyright:  © 2012 American Psychological Association

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription
Psychological Assessment

Psychometrical Assessment and Item Analysis of the General Health Questionnaire in Victims of Terrorism

David Delgado-Gomez, Jorge Lopez-Castroman, Victoria de Leon-Martinez, Enrique Baca-Garcia, Maria Luisa Cabanas-Arrate, Antonio Sanchez-Gonzalez, and David Aguado

Online First Publication, December 3, 2012. doi: 10.1037/a0030645

CITATION
Psychometrical Assessment and Item Analysis of the General Health Questionnaire in Victims of Terrorism

David Delgado-Gomez  
Universidad Carlos III

Jorge Lopez-Castroman  
Inserm, U1061 and IIS-Fundacion Jimenez Diaz

Victoria de Leon-Martinez  
IIS-Fundacion Jimenez Diaz

Enrique Baca-Garcia  
IIS-Fundacion Jimenez Diaz, New York State Psychiatric Institute, and Columbia University

Maria Luisa Cabanas-Arrate and Antonio Sanchez-Gonzalez  
Fundación Instituto de Victimología

David Aguado  
Instituto Ingeniería del Conocimiento

There is a need to assess the psychiatric morbidity that appears as a consequence of terrorist attacks. The General Health Questionnaire (GHQ) has been used to this end, but its psychometric properties have never been evaluated in a population affected by terrorism. A sample of 891 participants included 162 direct victims of terrorist attacks and 729 relatives of the victims. All participants were evaluated using the 28-item version of the GHQ (GHQ-28). We examined the reliability and external validity of scores on the scale using Cronbach’s alpha and Pearson correlation with the State-Trait Anxiety Inventory (STAI), respectively. The factor structure of the scale was analyzed with varimax rotation. Samejima’s (1969) graded response model was used to explore the item properties. The GHQ-28 scores showed good reliability and item-scale correlations. The factor analysis identified 3 factors: anxious-somatic symptoms, social dysfunction, and depression symptoms. All factors showed good correlation with the STAI. Before rotation, the first, second, and third factor explained 44.0%, 6.4%, and 5.0% of the variance, respectively. Varimax rotation redistributed the percentages of variance accounted for to 28.4%, 13.8%, and 13.2%, respectively. Items with the highest loadings in the first factor measured anxiety symptoms, whereas items with the highest loadings in the third factor measured suicide ideation. Samejima’s model found that high scores in suicide-related items were associated with severe depression. The factor structure of the GHQ-28 found in this study underscores the preeminence of anxiety symptoms among victims of terrorism and their relatives. Item response analysis identified the most difficult and significant items for each factor.

Keywords: factor analysis, item response theory, terrorism victims, GHQ-28

In 2010, 11,604 terrorist attacks killed 13,186 and injured 30,665 persons globally (National Counter Terrorism Center, 2011). The indirect consequences of these attacks on mental health may be more frequent (10–100 psychiatric casualties for every physical injury), and they have more significance, cost and long-term effects wise, than previously thought (Ursano & Friedman, 2006). The families of the victims, independently of their own exposure, may also present a high risk for the development of psychopathology (Hoven et al., 2009; Stoddard et al., 2011). The most frequent mental disorders in the aftermath of a terrorist attack are posttraumatic stress disorder (PTSD), depression, and other anxiety disorders (Henriksen, Bolton, & Sareen, 2010; Salguero, Cano-Vindel, Iruarrizaga, Fernandez-Berrocal, & Galea, 2011). The aftereffects of the trauma exposure can affect victims and their
relatives for years (Baca, Cabanas, & Baca-García, 2002). PTSD in particular often follows a chronic course in the general population (Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995) and after exposure to terrorist attacks (DiGrande, Neria, Brackbill, Pulliam, & Galea, 2011; Kawana, Ishimatu, & Kanda, 2001; North, Pfefferbaum, Kawasaki, Lee, & Spitznagel, 2011).

Although early detection of mental disorders in exposed populations may prevent future disabilities (Stoddard et al., 2011), the performance of screening instruments after a terrorist attack may present limitations. However, they have seldom been studied (Brewin, Fuchkan, Huntley, & Scragg, 2010). For instance, the diagnostic performance of screening instruments, such as the Trauma Screening Questionnaire, in an affected population was improved when the interval from the terrorist attacks was longer (Brewin et al., 2010). Moreover, there is a specific need for outreach to detect individuals with mental disorders after terrorist attacks (Brewin et al., 2008; Pfefferbaum, North, Flynn, Norris, & DeMartino, 2002).

One instrument that has been successfully used to assess the psychopathology of terrorism victims and their relatives is the General Health Questionnaire (GHQ; Baca et al., 2002; Boscarino, Figley, & Adams, 2004). The GHQ, in all its versions, is a self-administered Likert-type questionnaire frequently used to detect nonpsychotic mental disorders. The results of the GHQ might depend on the characteristics of the sample. For instance, the GHQ provided greater sensitivity and specificity in a clinical sample of 382 adult females than in a comparable sample of 154 males (Hobbs, Ballinger, Greenwood, Martin, & McClure, 1984). Differences were also observed among 260 primary care respondents: respondents younger than 40 years of age had 3 times more risk of being misclassified with the GHQ than respondents older than 40 years of age (de Jesus Mari & Williams, 1986). Moreover, Tarнопольский et al. indicated that validity coefficients could differ from one population to another. Therefore, it is necessary to analyze the reliability and external validity of the scores of the GHQ and its factor structure in any population under study (Tarnopolsky, Hand, McLean, Roberts, & Wiggins, 1979). To date, the GHQ has been validated in several populations varying in age (Baksheev, Robison, Cosgrave, Baker, & Yung, 2011; Costa et al., 2006), sex (Romans-Clarkson, Walton, Heribson, & Mullen, 1989; Tran, Tran, & Fisher, 2012), nationality (Politi, Piccinelli, & Wilkinson, 1994; Quek, Low, Razack, & Loh, 2001), and work status (Banks et al., 1980), however not in terrorism victims. Moreover, Tarнопольский et al. supported the validation of the GHQ on any particular population because validity coefficients could differ from one setting to another (Tarnopolsky et al., 1979).

The reliability, factor structure, and external validity represent the classical test theory in psychological assessment. This traditional assessment can be complemented with the item response theory (IRT), which is not commonly used to evaluate clinical instruments (Thomas, 2011) and, to our knowledge, has never been applied to the GHQ. The IRT analyses provide indices of discrimination and difficulty for each item, and therefore a deeper understanding of the questionnaire (P. J. Ferrando, 2001; Ploubidis & Frangou, 2011).

In this article, the psychometric properties of the GHQ as a screening questionnaire in a population of terrorism victims and their relatives were evaluated. Although the questionnaire has been used before to assess terrorism victims and their relatives (Baca et al., 2002), its psychometric properties have never been analyzed in this population. Thus, the first goal of this study was to analyze the reliability and external validity of scores of the 28-item version of the GHQ (GHQ-28) and its factor structure in this population. The second goal was to estimate the parameters of each item of the GHQ-28 using the IRT. The GHQ-28 items were then sorted according to the IRT. The item characteristic curves (ICCs) provided a visual understanding of the item responses. In conclusion, the IRT allowed a more robust estimate of the latent variable, given that the item parameters are independent of the sample.

Method

Participants

The sample was composed of 891 participants: 162 participants had experienced a terrorist attack, and 729 participants were relatives of the victims. The sample was obtained from interviews with victims of terrorist attacks in Spain and their immediate families (first-degree relatives). The average time from the terrorist attacks to the interviews was 13.5 years (SD = 6.6). This sample is part of the Phoenix Project (Baca et al., 2002), which was funded by the Association of Victims of Terrorism and designed to study the long-term consequences of terrorist attacks. The types of the terrorist attacks were mostly bombings (58.7%) or shootings (36.2%). All participants effectively completed the GHQ-28 and the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970).

The mean age of the sample was 38.7 ± 14.8 years. Although there was a majority of females in the total sample (60.6%), the victims were composed of mostly males (69.7% males). The victims and their relatives were mostly married (50.3%) or single (34.4%). Regarding the level of education in the sample, 28.2% reported primary studies or less, 46.0% reported secondary studies, and 25.8% reported university studies. A majority of the victims (74.9%) were policemen or army soldiers.

Measures

The GHQ-28. The GHQ-28 is a Likert-type scale composed of 28 items selected from the 60-item GHQ version by a principal component analysis (Goldberg & Hillier, 1979). Each item receives a discrete score ranging from 0 to 3. The 28 items were then divided into four groups. Items in each group measure a different trait. These four traits were somatic, anxiety and depression symptoms, and social dysfunction (see Table 2). Three different approaches have appeared in the literature to evaluate the questionnaire. They differ in the way the score of each item is calculated. In the classical GHQ approach, values 0 and 1 score a point value of 0, whereas values 2 and 3 score a point value of 1. In the Likert approach, values are not transformed. There also exists a third uncommon approach, called the c-GHQ, where the score of each item relies on whether the item is formulated positively or negatively (Richard, Lussier, Gagnon, & Lamarche, 2004). In this article, the Likert approach was used. It has been observed that the Likert method is the preferable method as it produces less skewed distributions and is more suitable for correlation analysis and intergroup comparisons (Banks et al., 1980). The Spanish version of the questionnaire was handed to the participants (Lobo & Muñoz, 1996).
The STAI. The STAI (Spielberger et al., 1970) is a Likert-type self-report questionnaire composed of 40 items. The first 20 items measure anxiety as a state, whereas the last 20 items measure anxiety as a trait with a single score for each. Previous literature has demonstrated a high reliability and validity for STAI scores in different samples (Quek, Low, Razack, Loh, & Chua, 2004; Zhang & Gao, 2012).

Procedure

Reliability, factor analysis, and external validity. The reliability of the GHQ-28 was estimated using Cronbach’s alpha. The adjusted correlations of each item with the GHQ-28 were also calculated. An exploratory factor analysis using varimax rotation was conducted to discover the structure of the GHQ-28. A maximum likelihood estimator was used for the exploratory factor analysis. Coefficients obtained with this method were similar to those obtained with other estimation methods, such as principal components and generalized least squares, except for an inversion of the second and third factor. However, the second and third factor presented very similar eigenvalues using any of the aforementioned methods (data not shown). In order to determine the factors to be retained, a parallel analysis was conducted following Hayton, Allen, and Scarpello (2004). The Bartlett method was used to calculate the factor scores. Finally, the external validity was analyzed in comparison with a questionnaire measuring anxiety (STAI) in agreement with Sánchez-López and Dresch (2008). We estimated the correlation between the STAI (trait and state variety (STAI) in agreement with Sánchez-López and Dresch (2008).

Removing Each Item in the GHQ-28

Adjusted Item-Scale Correlation and Cronbach’s Alphas After Removing Each Item in the GHQ-28

<table>
<thead>
<tr>
<th>Item</th>
<th>Adjusted item-scale correlation</th>
<th>Cronbach’s α if the item is eliminated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.603</td>
<td>0.954</td>
</tr>
<tr>
<td>2</td>
<td>0.647</td>
<td>0.954</td>
</tr>
<tr>
<td>3</td>
<td>0.766</td>
<td>0.953</td>
</tr>
<tr>
<td>4</td>
<td>0.684</td>
<td>0.953</td>
</tr>
<tr>
<td>5</td>
<td>0.589</td>
<td>0.954</td>
</tr>
<tr>
<td>6</td>
<td>0.649</td>
<td>0.954</td>
</tr>
<tr>
<td>7</td>
<td>0.615</td>
<td>0.954</td>
</tr>
<tr>
<td>8</td>
<td>0.673</td>
<td>0.954</td>
</tr>
<tr>
<td>9</td>
<td>0.657</td>
<td>0.954</td>
</tr>
<tr>
<td>10</td>
<td>0.774</td>
<td>0.953</td>
</tr>
<tr>
<td>11</td>
<td>0.769</td>
<td>0.953</td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>0.953</td>
</tr>
<tr>
<td>13</td>
<td>0.773</td>
<td>0.953</td>
</tr>
<tr>
<td>14</td>
<td>0.762</td>
<td>0.953</td>
</tr>
<tr>
<td>15</td>
<td>0.288</td>
<td>0.956</td>
</tr>
<tr>
<td>16</td>
<td>0.565</td>
<td>0.955</td>
</tr>
<tr>
<td>17</td>
<td>0.548</td>
<td>0.955</td>
</tr>
<tr>
<td>18</td>
<td>0.513</td>
<td>0.955</td>
</tr>
<tr>
<td>19</td>
<td>0.514</td>
<td>0.955</td>
</tr>
<tr>
<td>20</td>
<td>0.453</td>
<td>0.955</td>
</tr>
<tr>
<td>21</td>
<td>0.584</td>
<td>0.954</td>
</tr>
<tr>
<td>22</td>
<td>0.665</td>
<td>0.954</td>
</tr>
<tr>
<td>23</td>
<td>0.729</td>
<td>0.953</td>
</tr>
<tr>
<td>24</td>
<td>0.702</td>
<td>0.953</td>
</tr>
<tr>
<td>25</td>
<td>0.602</td>
<td>0.954</td>
</tr>
<tr>
<td>26</td>
<td>0.782</td>
<td>0.953</td>
</tr>
<tr>
<td>27</td>
<td>0.689</td>
<td>0.954</td>
</tr>
<tr>
<td>28</td>
<td>0.623</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Table 1

<table>
<thead>
<tr>
<th>Internal consistency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>α = 0.955</td>
</tr>
<tr>
<td></td>
<td>Std α = 0.954</td>
</tr>
<tr>
<td>Victims</td>
<td>α = 0.962</td>
</tr>
<tr>
<td></td>
<td>Std α = 0.961</td>
</tr>
<tr>
<td>Relatives</td>
<td>α = 0.951</td>
</tr>
<tr>
<td></td>
<td>Std α = 0.950</td>
</tr>
</tbody>
</table>

Note. Internal consistency of the GHQ-28 as measured with Cronbach’s alphas (Standardized) for the total sample (victims and their relatives). GHQ-28 = 28-item version of the General Health Questionnaire; Std = Standardized.
nation value, the easier the discrimination is between the distinct levels of the latent variable. These parameters are graphically represented in the ICCs. The ICCs represent the probability of each item response conditioned on the specific values of the latent variable. The slope of the curve of the location values determines the discrimination of an item. An example of ICC with three possible answers, and a discriminative parameter \( a_i \) of 2.5, and location parameters \( b_i \) of 11, 0, and 1 are displayed in Figure 2. In this article, the ICCs have been summarized in an expected average score for each item, in accordance with:

\[
\sum_{k=0}^{n} k \int_{-\infty}^{\infty} P(\kappa k \theta) f(\theta) \, d\theta,
\]

where \( f(\theta) \) is the density function of a standard Gaussian distribution. The average score allows us to sort the items by their inherent location for a given latent variable. For instance, a person with a higher latent variable would obtain higher scores in low-average score items than a person with a smaller latent variable. All procedures were performed using the ltm R statistical package.

<table>
<thead>
<tr>
<th>Factors</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue and percentage of explained variance before rotation</td>
<td>12.32</td>
<td>1.79</td>
<td>1.40</td>
</tr>
<tr>
<td>% Variance</td>
<td>43.99</td>
<td>6.39</td>
<td>5.01</td>
</tr>
<tr>
<td>% Cumulative variance</td>
<td>43.99</td>
<td>50.39</td>
<td>55.40</td>
</tr>
<tr>
<td>Eigenvalue and percentage of explained variance after rotation</td>
<td>7.95</td>
<td>3.85</td>
<td>3.70</td>
</tr>
<tr>
<td>% Variance</td>
<td>28.40</td>
<td>13.75</td>
<td>13.24</td>
</tr>
<tr>
<td>% Cumulative variance</td>
<td>28.40</td>
<td>42.16</td>
<td>55.40</td>
</tr>
</tbody>
</table>

Note. GHQ-28 = 28-item version of the General Health Questionnaire. Bold type identifies .40 or larger loads.

Results

Table 1 shows the Cronbach’s alphas (and the standardized alpha) for the total sample, the victims and their relatives together with the adjusted item-scale correlation, and the Cronbach’s alphas after removing each item. The calculated Cronbach’s alphas were superior to 0.95 in all cases. The adjusted item-scale correlation ranged from .45 to .78 for all the items, except Item 15 in which the item-scale correlation was .29.
Factor Analysis

Three factors were retained by the parallel analysis. They also showed an associated eigenvalue higher than 1 (see Figure 1). The eigenvalues of these factors together with the corresponding percentage of explained variance and the factor loadings are displayed in Table 2. The highest loadings of the first factor were found in the first 14 items (somatic and anxiety symptoms) and Item 26 (depression symptoms). The highest loadings of the second factor were located mainly in the items measuring social dysfunction symptoms. Finally, with the exception of Item 26, the highest loadings of the third factor were found in the items indicating depression symptoms. The loading of Item 22 was very similar for the first and third factor, but it was classified within the third factor in agreement with previous reports. When comparing the victims of terrorism and their relatives with regards to the factor scores, we found significant differences for the first ($p = .046$), second ($p < .001$), and third factor ($p = .028$). Factor structure and internal validity of samples separated by gender or class (victims or family members) showed no differences.

External validity. Table 3 shows the correlation of the GHQ-28 global score and the obtained factors with the state and trait global scores of anxiety according to the STAI. The first factor had the greatest correlation with the STAI resulting in a state score of 0.67 and a trait STAI score of 0.65.

IRT-based analyses. The factor analysis indicates that the GHQ-28 measures three dimensions or latent variables. We conducted an IRT-based analysis for each. On the basis of the factor analysis, items numbering from 1 to 14 and Item 26 were associated with the first latent variable, items numbering from 15 to 21 were associated with the second latent variable, and the remaining items were associated with the third latent variable. Table 4 shows the discrimination and location parameters for each item. The ICCs of the items with the highest and lowest expected average value for each dimension are displayed in Figure 2. We compared the average latent variables for each factor according to the IRT between the victims of terrorism and their relatives using $t$ tests. We found significant differences ($p < .001$) between the groups in the three factors.

Discussion

In this study, we analyzed the psychometric properties, including an item analysis using IRT, of the GHQ-28 in a sample of terrorism victims and their relatives. The reliability of the scores of GHQ-28 was high, as demonstrated by an alpha value of over 0.955 in the total sample (Bland & Altman, 1997). Moreover, with the exception of Item 15 (“Have you recently felt that you are playing a useful part in things?”), all items exhibited a reasonably good adjusted item-scale correlation with values ranging from .45 to .78. The factor structure and internal validity did not differ when separating the sample by gender or class (victims or family members). Previous literature supports a continuous increase of risk for mental disorders depending on the degree to which participants are affected by a terrorist act, independently of being directly affected or through a familiar bond (Baca, Cabanas, Perez-Rodriguez, &
Baca-Garcia, 2004). These reasons lead us to analyze victims of terrorism and their family members together.

The factor analysis identified three factors. The highest loadings of the first factor appeared in the group of anxiety and somatic symptoms. The highest loadings of the second and third factor were associated with the social dysfunction and depression items, respectively. Previous reports identified similar factors (Aderibigbe, Riley, Lewin, & Gureje, 1996; Gibbons, Flores de Arévalo, & Mónico, 2004; Vallejo, Jordan, Diaz, Comeche, & Ortega, 2007; Werneke, Goldberg, Yalcin, & Ustun, 2000). Notably, the highest loadings of the third factor correspond to symptoms of severe depression associated with suicidal behavior (see Items 24, 25, 27, and 28 in Table 2). This finding goes in line with the increased risk of suicide associated with suicidal behavior (see Items 24, 25, 27, and 28 in Table 2). Finally, the highly significant correlation between both the trait and state scores of the STAI and the three factors of the GHQ-28 supports the central role of anxiety symptoms in this population. Somatic symptoms, which are frequently associated with underlying anxiety (Romera et al., 2010; Zhu et al., 2012), also presented high loadings in the first factor.

We found that the five most discriminative items were Items 12, 7, 6, and 4, related with somatic symptoms of anxiety. On the contrary, Items 26, 28, both related with subjective feelings of anxiety, and Items 27, 28, both related with social dysfunction, associated with underlying anxiety (Romera et al., 2010; Zhu et al., 2012), also presented high loadings in the first factor. The IRT analysis demonstrates the importance of items within each identified factor. Interestingly, we found a certain gradient in the first factor that could reflect the severity of anxiety symptoms. We found that the five most discriminative items were Items 12 and 26, both related with subjective feelings of anxiety, and Items 7, 6, and 4, related with somatic symptoms of anxiety. On the contrary, the least discriminative items are those questions that address anxiety symptoms indirectly. For instance, Item 1 (“good health”), Item 10 (“under strain”), or Item 8 (“lost sleep”). Within the second factor (social dysfunction), the most discriminative items appeared to be closely related with the capacity to commit
oneself to any task (Item 15) and to making decisions (Item 20).
Finally, the most discriminative items among the depression symptoms seem to be the items regarding passive or active ideas of suicide (Items 25, 27, and 28). On the contrary, low self-esteem and hopelessness (Items 22 and 23), usual descriptors of depression, appeared to be the least discriminative.

The ICCs (see Figure 3) show that items included in the same factor may have very different discrimination powers. For instance, it can be observed that the latent variable in Item 1 has a wider range of values of location for each answer that can be chosen. This implies that small differences in the latent variable will not result in different response options. Thus, Item 1 is very discriminative and can provide accurate information on the actual value of the latent variable (the anxiety factor). Item 12 is the least discriminative item in this factor. Small changes in the value of the latent variable may change the participant’s response and therefore provides less valuable information. In the third factor, the small differences between the ICCs of the items indicate a similar

Figure 3. Item characteristic curves of the items with the lowest and highest expected average score of each factor.
capacity to discriminate the changes in the latent variable (depression).

This study presents several strengths. We examined a representative sample of victims of terrorism and their relatives. The traditional analyses of GHQ-28 were enhanced with a polotomic model of item response. However, the validity of our results could have been confirmed with other sources, such as a clinical assessment or other screening scales. We had no information of the diagnoses assigned to the participants in the sample; therefore, gender and age effects could not be analyzed regarding the diagnoses of mental disorders. The time gap between the terrorist attacks and the GHQ-28 assessment is another limitation. However, several studies have demonstrated the long-term effects of terrorist attacks on direct victims and their relatives (Baca et al., 2002; DiGrande et al., 2011; L. Ferrando et al., 2011; Kawana et al., 2001; North et al., 2011). It appears symptoms may be more stable 1 year after the traumatic event rather than in the immediate aftermath (Brewin, 2005).

The analysis of the GHQ-28 applied to this sample of victims of terrorism and their relatives demonstrated a high reliability and a good external validity, as well as a three-factor structure. The first factor had the highest loadings in anxiety symptoms, suggesting that anxiety could be the most frequent psychopathological finding in this population. A selection of the most discriminative items, based on the IRT, could be used to improve the screening for detection of mental disorders.

References


