

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

General Health Questionnaire (GHQ-12): the effect of negative items in its factorial structure

Valdiney V. Gouveia ¹
 Tiago Jessé Souza de Lima ¹
 Rildésia Silva Veloso Gouveia ²
 Leogildo Alves Freires ¹
 Larisse Helena Gomes Macedo Barbosa ¹

Abstract

The General Health Questionnaire (GHQ-12) detects non severe psychiatric illnesses. Although commonly treated as a general index, the definition of its factorial structure is cause for debate. This study aimed to test this structure comparing three models: two frequently cited in the literature (one- and bi-factorial) and a third, also one-factor, which controls response bias due to the negative wording of items. A total of 1,180 people participated (300 undergraduates, 311 military policemen, 274 elementary school teachers and 295 members of the general population) answering the GHQ-12 and demographic questions. Confirmatory factor analyses showed that the one-factor structure, controlling by the wording effect, gathered the best fit indexes, except among the military. This structure showed greater than 0.80 reliability in all groups. It was concluded that the one-factor model of the GHQ-12 is more appropriate, however, future studies are required with people with different occupations and levels of mental health.

Mental Health; Statistical Bias; Questionnaires; Validity of Tests

O Questionário de Saúde Geral (QSG), uma medida autoaplicável, do tipo lápis e papel, foi proposto originalmente por Goldberg & Williams ¹, composto por 60 itens, popularizando-se rapidamente ^{2,3}. Contribuíram para esse aspecto suas diversas versões abreviadas, formadas por 30, 28, 20 e 12 itens, produzidas baseando-se em estudos com análises fatoriais, que o tornou uma ferramenta útil em contextos que requerem avaliações rápidas sobre desconforto psicológico (por exemplo, clínica, organização, comunidade), sem comprometer evidências de confiabilidade de suas pontuações ^{4,5}.

A versão com 12 itens deste instrumento ¹ é a mais curta e, provavelmente, uma das mais empregadas para medir o bem-estar psicológico ⁶. Conta com o atrativo de ser de fácil aplicação, além de apresentar uma característica comum em medidas atitudinais: os itens são igualmente distribuídos entre os redigidos como positivos (e.g., “Tem se sentido capaz de tomar decisões?”) e negativos (e.g., “Tem pensado que você é uma pessoa que não serve para nada?”) ⁷. Essa característica, não obstante, pode não ser recomendável quando se tem por objetivo medir outros construtos, como traços de personalidade e indicadores de afetos e bem-estar subjetivo, podendo criar um fator espúrio ^{8,9}.

Comumente, o QSG-12 é avaliado como um índice unidimensional para mensurar morbidade psicológica ¹⁰. Entretanto, o debate acerca de sua estrutura fatorial parece longe de um con-

¹ Universidade Federal da Paraíba, João Pessoa, Brasil.

² Centro Universitário de João Pessoa, João Pessoa, Brasil.

Correspondência

V. V. Gouveia
 Universidade Federal da Paraíba.
 Rua Hortêncio Osterne
 Carneiro 598, João Pessoa, PB
 58035-120, Brasil.
 vvgouveia@gmail.com

senso. Diversos estudos apontam soluções fatoriais distintas, sendo as mais conhecidas aquelas com um, dois e três fatores^{11,12,13,14,15}. Porém, como Hankins¹⁶ indica, embora alguns estudos definam este instrumento como multifatorial, as evidências correspondentes podem ser proporcionadas por um fenômeno conhecido como efeito do método. Concretamente, o emprego de itens positivos e negativos em uma medida, embora possa minimizar o erro de aquiescência, produz potencialmente ao menos um fator adicional ou torna a estrutura fatorial mais complexa do que seria de fato⁹. Esse problema é o foco principal do presente trabalho, cujo objetivo é avaliar a estrutura fatorial do QSG-12, tendo em conta o possível efeito da redação dos itens.

No Brasil, foram encontrados ao menos cinco estudos que avaliaram a estrutura fatorial deste instrumento, considerando amostras da população geral¹⁷, jovens adolescentes¹⁸ e contextos laborais^{4,11,19}. Tais estudos empregaram análises exploratórias e/ou confirmatórias para testar sua dimensionalidade. Sarriera et al.¹⁸, utilizando análise de componentes principais (rotação *varimax*), encontraram três fatores: autoestima ($\alpha = 0,66$), depressão ($\alpha = 0,68$) e autoeficácia percebida ($\alpha = 0,54$), com consistência interna (α) de 0,80 para o total de itens. Borges & Argolo⁴ realizaram análise fatorial dos eixos principais (rotação *oblimin*), observando dois fatores: autoeficácia (itens 1, 3, 4, 6, 7, 8, 10, 11 e 12; $\alpha = 0,85$) e depressão e esgotamento emocional (itens 2, 6 e 9; $\alpha = 0,75$), admitindo um fator geral cujo α foi 0,80. Gouveia et al.¹⁷ testaram três modelos fatoriais, compostos por um, dois e três fatores, realizando análise fatorial confirmatória. Seus resultados indicaram o modelo bifatorial como o mais adequado, formado por depressão (itens 3, 4, 6, 8, 9, 10 e 11; $\alpha = 0,81$) e ansiedade (disfunção social) (itens 1, 2, 5 e 7; $\alpha = 0,66$). Oliveira¹⁹, empregando análise de componentes principais (*oblimin*), encontrou uma solução bifatorial: desconforto psicológico (itens 2, 5, 6, 7, 9, 10 e 11; $\alpha = 0,84$) e autoeficácia (itens 1, 3, 4, 8 e 12; $\alpha = 0,76$). Por fim, Gouveia et al.¹¹ realizaram análises fatoriais confirmatórias, testando e comparando os modelos uni e bifatorial desta medida. Os resultados apoiaram como mais adequada a estrutura bifatorial: depressão (itens 1, 2, 5, 6, 7 e 9; $\alpha = 0,85$) e disfunção social (itens 3, 4, 8, 10, 11 e 12; $\alpha = 0,82$); a solução unifatorial apresentou α de 0,89.

Com o fim de conhecer a realidade em outros países, realizou-se uma busca nas bases de dados PsycINFO e Google Acadêmico, em 18 de abril de 2011, utilizando como palavras-chave “GHQ-12” e “validity”. Na ocasião foram encontrados 15 artigos que tinham por objetivo principal testar a estrutura fatorial do QSG-12, dos quais três

empregaram abordagem estritamente exploratória, comumente com análise de componentes principais, quer admitindo rotação *oblimin*^{20,21} ou *promax*²². Montazeri et al.²¹, ao testarem a estrutura do QSG-12 em amostra iraniana, encontraram uma solução bidimensional: estresse psicológico (itens 1, 3, 4, 7, 8, 10 e 11) e disfunção social (itens 2, 5, 6, 9 e 12). Doi & Minowa²², em amostra japonesa, identificaram uma estrutura bifatorial para o grupo de mulheres, compreendendo os fatores estresse psicológico (itens 2, 5, 6, 9, 10 e 11) e disfunção social (itens 1, 3, 4, 7 e 8), porém trifatorial para os homens, representada pelos fatores estresse psicológico (itens 2, 5, 6, 9, 10 e 11), disfunção social (itens 1, 3, 4, 7 e 8) e felicidade (itens 3 e 12). López & Dresch²⁰, em contexto espanhol, observaram uma estrutura com três fatores: lidar com o sucesso (itens 1, 3, 4, 7, 8 e 12), autoestima (itens 6, 9, 10 e 11) e estresse (itens 2, 5 e 9). Destaca-se que nenhum desses trabalhos citou a consistência interna dos fatores específicos, mas apenas para a escala como um todo, que variaram de 0,76²⁰ a 0,87²¹. Essa diversidade de resultados é encarada por Ye²³ como decorrência do método empregado, isto é, análise de componentes principais, bem como dos critérios diferentes adotados pelos pesquisadores para a retenção dos fatores.

Partindo para uma abordagem confirmatória, encontram-se 12 estudos, dez dos quais a empregaram unicamente^{10,14,15,16,23,24,25,26,27,28} e outros dois a combinaram com análises exploratórias^{29,30}. Foram comumente testados os modelos uni (todos os itens saturando em um único fator: desconforto psicológico), bi (os itens divididos nos fatores ansiedade e depressão ou positivos e negativos, respectivamente) e trifatorial (ansiedade/depressão, disfunção social e perda da confiança; modelo de Graetz¹²). Desses estudos, quatro encontraram índices de ajustes mais favoráveis para o modelo trifatorial^{10,15,24,28}, três encontraram índices similares para os modelos bi e trifatorial^{14,26,29}, dois apontaram o modelo bifatorial como o mais ajustado^{27,30} e um sugeriu que o unifatorial era o que apresentava o melhor ajuste²⁵. Além desses, dois estudos^{16,23} testaram um modelo adicional, no qual o efeito presumível dos itens negativos foi controlado, encontrando índices de ajuste melhores nesta configuração.

Embora as soluções com dois e três fatores apresentem índices de ajuste adequados, algumas observações justificam avaliar o modelo que controla o efeito dos itens²³. No modelo bifatorial, por exemplo, os fatores são constituídos, em sua totalidade ou maioria (70% ou mais), por agrupamentos de itens que separam aqueles positivos dos negativos, evidenciando padrões diferentes de respostas possivelmente decorren-

tes de efeito metodológico. Evidências adicionais desse fenômeno podem ser observadas na literatura, como, por exemplo, em relação à estrutura fatorial da *Escala de Autoestima de Rosenberg* ³¹. Constituída por dez itens, sendo cinco negativos e cinco positivos, foi originalmente proposta para avaliar a autoestima global, embora investigações que empregaram análise fatorial exploratória tenham revelado uma estrutura bifatorial: autoestima positiva e autoestima negativa. No entanto, trabalhos recentes, que empregaram análises fatoriais confirmatórias para avaliar o efeito da redação dos itens, comprovam que esta estrutura se devia a tal efeito ³².

O emprego de itens escritos de formas negativa e positiva em instrumentos tem sido uma prática comum para evitar a tendência generalizada de concordar ou discordar com todos os itens, independentemente de seu conteúdo ^{33,34}. Como apontam Podsakoff et al. ³⁵, os itens negativos funcionam como lombadas cognitivas que requerem do respondente um processo cognitivo mais elaborado, ao invés de automático. Contudo, alguns estudos empíricos que buscaram avaliar as propriedades psicométricas de tais instrumentos, sugerem que a prática de empregar itens positivos e negativos pode introduzir erros sistemáticos de medida que afetam as análises e a interpretação dos resultados ^{33,36}. Esse tipo de erro ou viés de resposta representa o efeito do método derivado das características dos itens ³⁵, isto é, traduz determinada variância sistematicamente introduzida nos resultados decorrentes do método empregado para mensurar o traço ou construto investigado.

Uma das explicações possíveis para o fenômeno de itens negativos e positivos formarem fatores separados diz respeito à dificuldade de os respondentes processarem cognitivamente itens escritos de maneira inversa. Segundo Schmitt & Stults ³⁷, isso ocorre devido à falta de atenção, variação na educação ou aversão a conteúdos emocionais negativos. Desse modo, esse erro potencial de resposta introduz variação que não está associada ao construto, sendo de caráter exclusivo dos itens negativos, uma vez que afeta a medida em parte, diferentemente do erro randômico de mensuração, que atinge toda a sua extensão.

O problema anteriormente citado parece se restringir às estruturas uni e bifatorial. Não obstante, é preciso entender que o modelo bifatorial é aninhado com o trifatorial, isto é, um terceiro fator é, em realidade, o desmembramento dos itens 10 e 11 que abarcam conteúdos negativos (ansiedade/depressão), formando o fator perda de confiança ¹². Entretanto, ressalta-se, nos quatro estudos em que o modelo trifatorial se mos-

trou mais adequado ^{10,15,24,28}, os dois fatores previamente identificados se correlacionaram fortemente entre eles, com correlação média de 0,89, variando de 0,84 ¹⁵ a 0,93 ²⁴. Desse modo, embora se argumente que há um ganho adicional de informação ao considerar três fatores, ao menos estatisticamente pode ser difícil diferenciá-los.

Em resumo, não parece haver consenso quanto à estrutura fatorial do QSG-12, quer no Brasil ou em outros países, embora a maior parte das pesquisas tenha encontrado uma solução bifatorial. Em razão das evidências apresentadas, o presente estudo foi conduzido com o objetivo de avaliar três modelos fatoriais alternativos, empregando-se abordagem confirmatória: (a) unifatorial, representando um fator geral de desconforto psicológico; (b) bifatorial, dividindo os itens em positivos e negativos e; finalmente, (c) um terceiro também unifatorial, porém controlando o efeito dos itens negativos.

Método

Participantes

Participaram do presente trabalho 1.180 pessoas. A amostra foi de conveniência (não probabilística), ou seja, foi composta por pessoas que anuíram ao serem solicitadas a colaborar com o estudo. Desse total, 300 eram estudantes universitários, com idade média de 24 anos (desvio padrão – DP = 5,80), a maioria do sexo feminino (58,3%) e solteira (82,7%); 311 eram policiais militares, com idade média de 36 anos (DP = 1,92), a maioria do sexo masculino (94,5%) e casada (76,2%); 274 professores dos ensinos fundamental e médio, com idade média de 33,5 anos (DP = 9,00), sendo a maioria do sexo feminino (92%) e solteira (52%); e, por fim, 295 compreendiam pessoas da população em geral, com idade média de 34,2 anos (DP = 13,60), majoritariamente do sexo masculino (51%) e solteira (46%). Tomando a amostra total, os participantes foram distribuídos equitativamente quanto ao sexo, com idade média de 31,9 anos (DP = 10,20).

Instrumentos

Os participantes responderam ao QSG-12. Esse instrumento é uma versão reduzida do QSG-60, de Goldberg & Williams ¹. O QSG-12 é composto por 12 itens que avaliam o quanto a pessoa tem experimentado os sintomas descritos, devendo ser dadas as respostas em escala de quatro pontos. No caso de itens negativos (por exemplo, “Tem se sentido pouco feliz e deprimido”), as alternativas de resposta variam de 1 = absolu-

tamente não, a 4 = muito mais que de costume; em caso de itens positivos (por exemplo, “Tem se sentido capaz de tomar decisões?”), as respostas variaram de 1 = mais que de costume, a 4 = muito menos que de costume. Nesse sentido, os itens negativos foram invertidos, sendo a menor pontuação indicativa de melhor nível de bem-estar psicológico. No Brasil, o QSG-12 tem apresentado evidências satisfatórias de validade e precisão^{4,17,18}, embora, como apontado anteriormente, não se tenha uma estrutura fatorial consensual. A consistência interna global da escala tem se mostrado adequada, com alfas de Cronbach comumente acima de 0,80.

Além desse instrumento, foram incluídas perguntas para caracterização da amostra (sexo, idade, estado civil e ocupação profissional).

Procedimento

Procedemos à coleta individualmente, ainda que em ambientes coletivos, em seus respectivos locais de trabalho ou estudo. A aplicação foi levada a cabo por colaboradores devidamente treinados para este propósito. Ao abordar as pessoas, os aplicadores solicitavam sua colaboração no sentido de participar em uma pesquisa, de acordo com as seguintes instruções: (a) comentavam que estava sendo realizado um estudo para conhecer fatores ligados à sua saúde geral; (b) enfatizavam a necessidade de que as respostas fossem dadas individualmente, de acordo com o que se passava com cada uma; e (c) indicavam que toda a informação era confidencial, que somente seriam tratadas estatisticamente de forma conjunta, garantindo-lhes o anonimato e o sigilo de suas respostas, de acordo com a *Resolução nº. 196/96* do Conselho Nacional de Saúde. Quando necessária, era solicitada a anuência da direção da empresa/organização na qual os participantes se encontravam. Em média, 10 minutos foram suficientes para concluir cada participação.

Análise de dados

Os dados foram analisados com o PASW e o AMOS, ambos em suas versões 18 (SPSS Inc., Chicago, Estados Unidos). Com o objetivo de comparar os modelos fatoriais do QSG-12, foram realizadas múltiplas análises fatoriais confirmatórias, considerando a matriz de covariância e empregando-se o método de estimação *ML (Maximum Likelihood)*. O ajuste dos modelos foi verificado por meio dos seguintes indicadores^{38,39}: a) χ^2 (qui-quadrado) – testa a probabilidade de o modelo teórico se ajustar aos dados; quanto maior este valor, pior o ajustamento. Por esse ser sensível ao tamanho da amostra (amostras gran-

des, isto é, $n > 200$ podem favorecer o descarte de um modelo adequado), deve-se interpretá-lo com cautela, valendo-se de sua razão em relação aos graus de liberdade correspondentes ($\chi^2/g.l.$). Esse valor deve ficar entre 2 e 3, preferencialmente, admitindo-se até 5. Contudo, a prática sugere que poderá padecer de viés quando a amostra for superior a 1.000 participantes.

b) AGFI (*Adjusted Goodness-of-Fit Index*) – indica a proporção de variância-covariância nos dados explicada pelo modelo, tendo em conta os graus de liberdade e o número de variáveis consideradas. Varia de 0 (nenhum ajuste) a 1 (ajuste pleno), com valores na casa de 0,90 (ou superiores) indicando um ajuste satisfatório.

c) CFI (*Comparative Fit Index*) – é um índice comparativo de ajuste do modelo, que permite comparar a matriz de covariância predita pelo modelo com aquela realmente observada. São admitidos como expressando um ajuste adequado os valores de CFI na casa de 0,90 ou superiores.

d) RMSEA (*Root-Mean-Square Error of Approximation*) – é considerado um indicador de “maldade” de ajuste, isto é, valores altos sugerem um modelo não ajustado. Assume-se como ideal que a RMSEA se situe entre 0,05 e 0,08, aceitando-se valores até 0,10, desde que caia no intervalo de 90% de confiança (IC90%).

Com o objetivo de comparar os modelos alternativos, consideraram-se os seguintes indicadores: $\Delta\chi^2$, CAIC (*Consistent Akaike Information Criterion*) e ECVI (*Expected Cross Validation Index*). Diferença estatisticamente significativa do $\Delta\chi^2$, penalizando o modelo com maior χ^2 , e valores menores de CAIC e ECVI definem um modelo mais adequado.

Resultados

Inicialmente, calcularam-se as médias, os desvios padrões e as correlações entre os 12 itens do QSG. Esses valores, conjuntamente com a assimetria e a curtose, são apresentados na Tabela 1. Como se pode observar, a média dos itens positivos (itens 1, 3, 4, 7, 8 e 12) foi de 2,18 (DP = 0,69), variando de 2,00 (item 3) a 2,35 (item 7), enquanto que a média dos itens negativos (itens 2, 5, 6, 9, 10 e 11) foi de 2,13 (DP = 0,87), variando de 1,86 (item 11) a 2,36 (item 5). Essas médias foram estatisticamente diferentes ($t = 3,10$; valor de $p < 0,01$).

Os dois grupos de itens também apresentaram padrões diferentes de curtose, com os negativos tendo valores maiores ($m = -1,12$; DP = 0,10) do que os positivos ($m = -0,57$; DP = 0,06) ($t = 11,28$; valor de $p < 0,05$), sugerindo que os primeiros desviam mais de uma distribuição

Tabela 1

Estatísticas descritivas e correlações para os itens do *Questionário de Saúde Geral* (QSG-12).

Itens	1	2 *	3	4	5 *	6 *	7	8	9 *	10 *	11 *	12
2 *	-0,27	1,00										
3	0,31	-0,34	1,00									
4	0,37	-0,27	0,65	1,00								
5 *	-0,38	0,51	-0,25	0,25	1,00							
6 *	-0,36	0,45	-0,51	-0,52	0,47	1,00						
7	0,43	-0,31	0,39	0,41	-0,44	-0,45	1,00					
8	0,36	-0,34	0,59	0,64	-0,30	-0,53	0,48	1,00				
9 *	-0,40	0,51	-0,40	-0,48	0,57	0,60	-0,44	-0,50	1,00			
10 *	-0,36	0,39	-0,66	-0,64	0,38	0,62	-0,47	-0,63	0,64	1,00		
11 *	-0,30	0,37	-0,72	-0,66	0,29	0,59	-0,42	-0,59	0,57	0,81	1,00	
12	0,34	-0,34	0,62	0,61	-0,30	-0,51	0,47	0,63	-0,54	-0,62	-0,64	1,00
m	2,34	2,26	2,00	2,08	2,36	2,19	2,35	2,18	2,11	1,98	1,86	2,11
DP	0,81	1,07	0,99	0,97	1,01	1,11	0,82	0,92	1,08	1,21	1,25	0,93
Variância	0,66	1,14	0,99	0,94	1,03	1,24	0,68	0,84	1,17	1,47	1,56	0,87
Assimetria	0,118	0,283	0,683	0,596	0,091	0,408	0,184	0,416	0,446	0,732	0,940	0,515
Curtose	-0,50	-1,18	-0,61	-0,61	-1,12	-1,21	-0,48	-0,62	-1,15	-1,12	-0,92	-0,58

* Item negativo.

DP: desvio padrão; m: média.

normal, apresentando uma curva mais achatada (platicúrtica). Em relação à assimetria, os dois conjuntos de itens não se diferenciaram entre si ($t = -0,41$; valor de $p > 0,05$), tendo os itens positivos apresentado média de $-0,41$ ($DP = 0,23$) e os negativos de $0,48$ ($DP = 0,31$). Adicionalmente, computaram-se as variâncias dos itens. Comprova-se na Tabela 1 que a variância dos itens negativos ($m = 1,27$; $DP = 0,20$) foi maior do que a dos positivos ($m = 0,83$; $DP = 0,13$) ($t = -4,38$; valor de $p < 0,01$). Portanto, os itens negativos apresentaram maior variabilidade de respostas.

Em seguida, procurou-se comparar o ajuste dos três modelos por meio de análises fatoriais confirmatórias. Foi testado, além dos modelos clássicos, unifatorial e bifatorial, um modelo unifatorial alternativo no qual se controlou o efeito presumível de viés de resposta em razão de itens negativos. Nesse modelo, entende-se que os vieses provocados por tais itens podem afetar diretamente a estrutura fatorial da medida, criando ao menos um fator artificial (espúrio). Portanto, esse terceiro modelo inclui a possibilidade de lidar com um fator latente denominado viés de resposta devido à redação negativa dos itens. De acordo com a Figura 1, para fins de análise foi definido um fator latente que reuniu unicamente os itens negativos, de modo que o programa pudesse computar a influência desse fator espúrio, decorrente da redação dos itens

negativos, sobre a estrutura fatorial. As análises foram realizadas separadamente para as quatro subamostras, sendo os resultados apresentados na Tabela 2.

Conforme descrito nessa tabela, os resultados apontam que, nas amostras de estudantes, população em geral e professores, o modelo três (um fator geral, controlando o efeito da redação negativa dos itens) apresentou índices de ajuste melhores do que os modelos com dois e, principalmente, um fator. Apenas na amostra de militares os modelos três e dois não apresentaram diferença significativa de seus respectivos qui-quadrados. Finalmente, os alfas de Cronbach para a estrutura unifatorial, segundo a subamostra, foram como seguem: 0,83 (professores), 0,84 (policiais militares), 0,84 (população em geral) e 0,86 (estudantes).

A estrutura unifatorial, controlando o efeito dos itens redigidos negativamente, é representada na Figura 1. É importante destacar que todas as saturações (os Lambdas, λ) foram estatisticamente diferentes de zero ($\lambda \neq 0$; $z > 1,96$, $p < 0,05$), indicando que cada um dos 12 itens desta versão abreviada do QSG compreende um marcador importante do construto desconforto psicológico.

Figura 1

Estrutura fatorial controlando o efeito dos itens. Questionário de Saúde Geral (QSG-12).

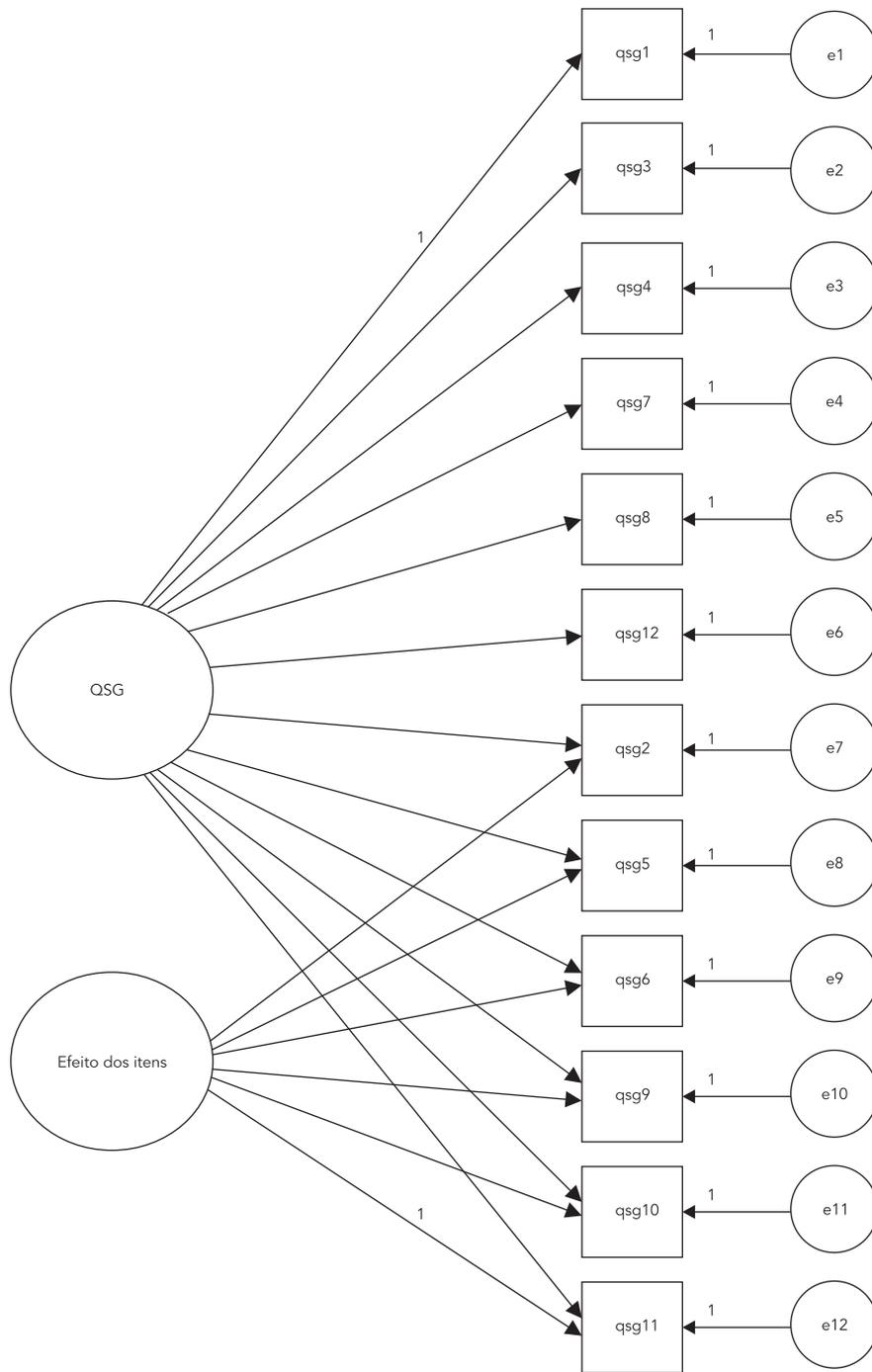


Tabela 2

Índices de ajuste dos modelos fatoriais alternativos do *Questionário de Saúde Geral* (QSG-12).

Índices		χ^2	gl	χ^2/gl	AGFI	CFI	RMSEA (IC90%)	CAIC	ECVI	$\Delta\chi^2(\text{gl})$
Estudante	1	256,68	54	4,75	0,79	0,82	0,112 (0,098-0,126)	417,58	1,02	86,74 (6) *
	2	199,35	53	3,76	0,82	0,87	0,096 (0,082-0,110)	366,95	0,834	29,41 (5) *
	3	169,94	48	3,54	0,85	0,89	0,092 (0,077-0,107)	371,05	0,769	-
Militares	1	210,99	54	3,91	0,83	0,85	0,097 (0,083-0,111)	372,74	0,835	95,20 (6) *
	2	121,40	53	2,29	0,90	0,93	0,065 (0,049- 0,080)	289,89	0,553	5,61 (5)
	3	115,79	48	2,41	0,90	0,93	0,067 (0,052-0,083)	317,98	0,567	-
População em geral	1	274,16	54	5,08	0,78	0,78	0,118 (0,104-0,132)	434,65	1,09	108,71 (6) *
	2	216,59	53	4,09	0,82	0,83	0,102 (0,088-0,117)	383,76	0,907	51,14 (5) *
	3	165,45	48	3,45	0,85	0,88	0,091 (0,076-0,107)	366,06	0,767	-
Professores	1	168,67	54	3,12	0,86	0,85	0,088 (0,073-0,103)	327,39	0,794	47,48 (6) *
	2	136,79	53	2,58	0,88	0,89	0,076 (0,061-0,092)	302,12	0,684	15,60 (5) *
	3	121,19	48	2,52	0,88	0,90	0,075 (0,058-0,091)	319,59	0,664	-

1: modelo unifatorial; 2: modelo bifatorial; 3: modelo controlando o efeito dos itens; AGFI: *Adjusted Goodness-of-Fit Index*; CAIC: *Consistent Akaike information Criterion*; CFI: *Comparative Fit Index*; ECVI: *Expected Cross Validation Index*; IC90%: intervalo de 90% de confiança; gl: graus de liberdade; RMSEA: *Root-Mean-Square Error of Approximation*.

* $p < 0,05$.

Discussão

O presente estudo teve como objetivo principal testar a estrutura fatorial do QSG-12, comparando três modelos, dos quais dois são comumente reportados na literatura (uni e bifatorial) e um terceiro, também unifatorial, que controla o efeito metodológico associado com a redação de itens negativos^{23,40}. Confia-se que os achados reportados contribuam para a discussão acerca da estrutura fatorial deste instrumento, oferecendo uma explicação alternativa para a inconsistência entre os resultados de estudos prévios. Embora tenham-se considerado quatro grupos diferentes, não foi o propósito compará-los em função de sua saúde mental, muito menos generalizar os resultados obtidos. Pretendeu-se unicamente assegurar diversidade amostral, favorecendo um teste amplo de hipótese de um fator de viés de resposta. Nesse sentido, o presente trabalho compreende um empreendimento de natureza psicométrica, buscando avaliar o efeito, na estrutura fatorial, de contar com itens redigidos de formas positiva e negativa.

A matriz de correlações interitens apresentou um padrão no qual ambos os grupos de itens, positivos e negativos, correlacionam-se positivamente entre eles, e negativamente com aqueles do grupo oposto. Ao analisar as variâncias dos itens, percebeu-se que os negativos apresentaram maior variabilidade do que os positivos, além de terem se diferenciado em relação à cur-

tose, apresentando uma curva mais platicúrtica. Portanto, inspecionando as correlações e variâncias, pareceu emergir uma estrutura com dois ou mais fatores. Contudo, esses padrões não corresponderam a construtos diferentes, resultando, provavelmente, da forma como os itens foram redigidos. Conforme exposto anteriormente, alguns fatores podem explicar esse fenômeno, como, por exemplo, a dificuldade de os respondentes processarem itens escritos de maneira inversa ou falta de atenção ao responder, não notando que a escala de resposta mudou. Como adverte Hankins¹⁶, o erro randômico da medida é aproximadamente igual entre todos os itens de um instrumento, assim qualquer diferença na variância entre itens positivos e negativos pode ser atribuída ao viés de resposta associado com a redação destes últimos.

As análises fatoriais confirmatórias apontaram que o modelo três (unifatorial, controlando o viés de resposta devido à redação dos itens negativos) apresentou índices de ajuste que, embora não sejam excepcionais, situaram-se próximos aos recomendados na literatura^{38,39}. Esses também foram coerentes com os achados de Gouveia et al.¹¹, os quais testaram os modelos uni e bifatorial em amostra de médicos de todo o Brasil, porém sem impor a restrição do fator latente de viés de resposta. Na ocasião, seu modelo bifatorial foi mais adequado, porém, refletiu, predominantemente, a dicotomia entre itens positivos (*e.g.*, 3, 4, 8 e 10) e negativos (*e.g.*,

2, 5, 7 e 9) desta medida. A propósito da alternativa do modelo bifatorial, Ye²³ aponta que os fatores respectivos costumam emergir repetidamente na literatura como forte e inversamente correlacionados entre si, além de apresentarem padrões de correlação similares com critérios externos. Embora seja possível argumentar que ao considerar mais fatores para o QSG-12 tenha-se informações adicionais, na prática isto pode não ter ocorrido aqui; concretamente, não se vislumbrou qualquer distinção substancial entre os fatores com itens negativos e positivos.

Por fim, diante das evidências e tendo em conta a estrutura fatorial identificada como mais adequada nas múltiplas amostras, recomenda-se que o QSG-12 seja considerado como uma medida unidimensional, provendo assim uma representação mais parcimoniosa do construto desconforto psicológico. Essa conclusão é consistente com a que oferece Ye²³, que indica que tal medida compreende uma ferramenta para avaliação geral da saúde mental, sendo preferível uma estrutura unidimensional. Shevlin & Adamson¹⁵, procurando avaliar o poder preditivo dos modelos com um e três fatores²¹ em relação a outras variáveis (ansiedade, stress e suporte social), observaram que o unifatorial (fator de primeira ordem) explicou uma proporção moderada de suas variâncias; já o modelo com três fatores, controlando o efeito do fator de primeira ordem, não apresentou qualquer coeficiente de regressão significativo em relação a tais variáveis.

Esses autores concluem que, com base nos valores de R², há pouca evidência do poder explicativo útil de se empregar o QSG-12 como uma medida multifatorial, afirmando que assumir uma pontuação total não resultará em perda substancial de informações. Contar com um fator único, no âmbito da pesquisa, torna mais parcimonioso conhecer os correlatos da saúde mental; no caso de aplicabilidade, parece bem mais direta a avaliação clínica pautada em um fator, contando com um ponto de corte único.

Não obstante, sugere-se cautela quanto à generalização dos achados, uma vez que as amostras aqui tratadas foram obtidas por conveniência, não se constituindo em estratos probabilísticos da população-meta. Além disso, recomenda-se que estudos futuros avaliem o efeito de variáveis como habilidade de leitura e nível de concentração dos participantes no momento de responderem aos instrumentos com itens positivos e negativos. Alternativamente, no caso específico do instrumento em pauta, poder-se-ia checar se a estrutura unifatorial é mesmo suficiente em razão do tipo de ocupação profissional ou grau de saúde mental; sabe-se que os policiais apresentam frequência elevada de sofrimento psíquico⁴¹, o que pode fazer com que tenham lugar fatores diferentes para retratar sintomas ou distúrbios específicos na medida considerada, mesmo depois de controlar o viés de redação dos itens.

Resumo

O Questionário de Saúde Geral (QSG-12) detecta doenças psiquiátricas não severas. Embora comumente tratado como um índice geral, a definição de sua estrutura fatorial suscita debates. Este trabalho objetivou testar tal estrutura, comparando três modelos: dois frequentemente citados na literatura (uni e bifatorial) e um terceiro, também unifatorial, que controla o viés de resposta devido à redação dos itens. Participaram 1.180 pessoas (300 estudantes universitários; 311 policiais militares; 274 professores do ensino fundamental; e 295 membros da população em geral), que responderam ao QSG-12 e perguntas demográficas. Análises

fatoriais confirmatórias apontaram que a estrutura unifatorial, controlando o efeito dos itens negativos, reuniu os melhores índices de ajuste, excetuando entre os militares. Essa estrutura apresentou consistência interna superior a 0,80 em todos os grupos. Concluiu-se que o QSG-12 é mais adequado como unifatorial, embora se indique a necessidade de estudos futuros com pessoas de profissões e níveis de saúde mental diferentes.

Saúde Mental; Viés Estatístico; Questionários; Validade dos Testes

Colaboradores

V. V. Gouveia contribuiu na elaboração da introdução, análise e discussão dos dados. T. J. S. Lima colaborou na elaboração da introdução, análise dos dados e discussão dos resultados. R. S. V. Gouveia participou na discussão dos resultados e revisão final do manuscrito. L. A. Freires contribuiu revisando criticamente o conteúdo teórico, apontando direcionamentos futuros e limitações do estudo, além de colaborar na coleta de dados. L. H. G. M. Barbosa colaborou na coleta de dados, preparação de tabelas e revisão do texto.

Referências

1. Goldberg DP, Williams PA. A user's guide to the General Health Questionnaire. Windsor: NFER-Nelson; 1988.
2. Carvalho HW, Patrick CJ, Jorge MR, Andreoli SB. Validation of the structural coherency of the General Health Questionnaire. *Rev Bras Psiquiatr* 2011; 33:59-63.
3. Pasquali L, Gouveia VV, Andriola WB, Miranda FJ, Ramos ALM. Questionário de Saúde Geral de Goldberg (QSG): adaptação brasileira. *Psicol Teor Pesqui* 1994; 10:421-37.
4. Borges LO, Argolo JCT. Adaptação e validação de uma escala de bem-estar psicológico para uso em estudos ocupacionais. *Aval Psicol* 2002; 1:17-27.
5. Kawada T, Otsuka T, Inagaki H, Wakayama Y, Katsumata M, Li Q, et al. Relationship among lifestyles, aging and psychological wellbeing using the General Health Questionnaire 12-items in Japanese working men. *Aging Male* 2011; 14:115-18.
6. Chipimo PJ, Fylkesnes K. Comparative validity of screening instruments for mental distress in Zambia. *Clin Pract Epidemiol Ment Health* 2010; 6:4-15.
7. Stewart TJ, Frye AW. Investigating the use of negatively phrased survey items in medical education settings: Common wisdom or common mistake? *Acad Med* 2004; 79:18-20.
8. Ryff CD, Keyes CLM. The structure of psychological well-being revisited. *J Pers Soc Psychol* 1995; 69:719-27.
9. Spector PE, Katwyk PTV, Brannick MT, Chen PY. When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *J Manage* 1997; 23:659-77.
10. Gao F, Luo N, Thumboo J, Fones C, Li S, Cheung Y. Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health Qual Life Outcomes* 2004; 2:63.
11. Gouveia VV, Barbosa GA, Andrade EO, Carneiro MB. Factorial validity and reliability of the General Health Questionnaire (GHQ-12) in the Brazilian physician population. *Cad Saúde Pública* 2010; 26:1439-45.
12. Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol* 1991; 26:132-8.

13. Martin AJ. Assessing the multidimensionality of the 12-item General Health Questionnaire. *Psychol Rep* 1999; 84:927-35.
14. Salama-Younes M, Montazeri A, Ismaïl A, Roncin C. 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 Qual Life Outcomes* 2009; 7:22.
15. Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychol Assess* 2005; 17:231-6.
16. Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clin Pract Epidemiol Ment Health* 2008; 4:10.
17. Gouveia VV, Chaves SS, Oliveira IC, Dias MR, Gouveia RSV, Andrade PR. A utilização do QSG-12 na população geral: estudo de sua validade de construto. *Teoria e Pesquisa* 2003; 19:241-8.
18. Sarriera JC, Schwarcz C, Câmara SG. Bem-estar psicológico: análise fatorial da escala de Goldberg (GHQ-12) numa amostra de jovens. *Psicol Reflex Crit* 1996; 9:293-306.
19. Oliveira GE. Trabalho e bem-estar subjetivo: compreendendo a situação laboral dos médicos [Tese de Doutorado]. João Pessoa: Universidade Federal da Paraíba; 2008.
20. López MPS, Dresch V. The 12-item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. *Psicothema* 2008; 20:839-43.
21. Montazeri A, Harirchi AM, Shariati M, Garmaroudi G, Ebadi M, Fateh A. The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. *Health Qual Life Outcomes* 2003; 1:66.
22. Doi Y, Minowa M. Factor structure of the 12-item general health questionnaire in the Japanese general adult population. *Psychiatry Clin Neurosci* 2003; 57:379-83.
23. Ye S. Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences* 2009; 46:191-201.
24. French D, Tait RJ. Measurement invariance in the general health questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004; 13: 1-7.
25. González M, Ibáñez I. Cuestionario de Salud General (GHQ-12): Comparación de dos modelos factoriales. *Psiquatria.com* 2001; 5:1.
26. Ip WY, Martin CR. Psychometric properties of the 12-item General Health Questionnaire (GHQ-12) in Chinese women during pregnancy and in the postnatal period. *Psychol Health Med* 2007; 11:60-9.
27. Kalliath TJ, Driscoll MP, Brough P. A confirmatory factor analysis of the General Health Questionnaire-12. *Stress Health* 2004; 20:11-20.
28. Mäkikangas A, Feldt T, Kinnunen U, Tolvanen A, Kinnunen ML, Pulkkinen L. The factor structure and factorial invariance of the 12-item general health questionnaire (GHQ-12) across time: Evidence from two community based samples. *Psychol Assess* 2006; 18:444-51.
29. Ip WY, Martin CR. Factor structure of the Chinese version of the 12-item General Health Questionnaire (GHQ-12) in pregnancy. *J Reprod Infant Psychol* 2006; 24:88-98.
30. Hu Y, Brown SS, Twigg L, Weich S. Can the 12-item General Health Questionnaire be used to measure positive mental health? *Psychol Med* 2007; 37:1006-13.
31. Rosenberg M. *Society and the adolescent self-image*. Middletown: Wesleyan University Press; 1989.
32. Greenberger E, Chen C, Dmitrieva J, Farruggia SP. Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter? *Pers Individ Dif* 2003; 35:1241-54.
33. Distefano C, Motl, RW. Further investigating method effects associated with negatively worded items on self-report surveys. *Struct Equ Modeling* 2006; 13:440-64.
34. Cronbach LJ. Further evidence on response sets and test design. *Educ Psychol Meas* 1950; 10:3-31.
35. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff MP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol* 2003; 88:879-903.
36. Chen YH, Gobiolf GR, Dedrick RE. Detecting effects of positively and negatively worded items on a self-concept scale for third and sixth grade elementary students. In: *Annual Meeting of the Florida Educational Research Association*. Tampa: Florida Educational Research Association; 2007. p. 30.
37. Schmitt N, Stults DM. Factors defined by negatively keyed items: the results of careless respondents? *Appl Psychol Meas* 1985; 9:367-73
38. Byrne BM. *Structural equation modeling with Amos: basic concepts, applications, and programming*. 2th Ed. New York: Routledge; 2010.
39. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 5th Ed. Boston: Allyn & Bacon; 2007.
40. Woods CM. Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *J Psychopathol Behav Assess* 2006; 28:189-94.
41. Souza ER, Franco LG, Meireles CC, Ferreira VT, Santos NC. Sofrimento psíquico entre policiais civis: uma análise sob a ótica de gênero. *Cad Saúde Pública* 2007; 23:105-14.

Recebido em 08/Jun/2011

Versão final reapresentada em 13/Set/2011

Aprovado em 03/Out/2011