Original article

A brief homophobia scale in medical students from two universities: Results of a refinement process ☆

Adalberto Campo-Arias [a,b,]*, *Edwin Herazo* [a], *Heidi Celina Oviedo* [a,c]

[a] *Instituto de Investigación del Comportamiento Humano (Human Behavioral Research Institute), Bogotá, Colombia*
[b] *Programa de Medicina, Facultad de Ciencias de la Salud, Universidad del Magdalena, Santa Marta, Colombia*
[c] *Universidad Autónoma de Bucaramanga, Colombia*

A B S T R A C T

Background: The process of evaluating measurement scales is an ongoing procedure that requires revisions and adaptations according to the characteristics of the participants. The Homophobia Scale of seven items (EHF-7) has showed acceptable performance in medical students attending to two universities in Colombia. However, performance of some items was poor and could be removed, with an improvement in the psychometric findings of items retained.

Objective: To review the psychometric functioning and refine the content of EHF-7 among medical students from two Colombian universities.

Methods: A group of 667 students from the first to tenth semester participated in the research. Theirs ages were between 18 and 34 (mean, 20.9 ± 2.7) years-old, and 60.6% were females. Cronbach alpha (α) and omega of McDonald (Ω) were calculated as indicators of reliability and to refine the scale, an exploratory (EFA) and confirmatory factor analysis (CFA) was performed.

Results: EHF-7 showed $\alpha = .793$ and $\Omega = .796$ and a main factor that explained 45.2% of the total variance. EFA and CFA suggested the suppression of three items. The four-item version (EHF-4) reached an $\alpha = .770$ and $\Omega = .775$, with a single factor that accounted for 59.7% of the total variance. CFA showed better indexes ($\chi^2 = 3.622$; $df = 1$; $p = 0.057$; Root-mean-square error of approximation (RMSEA) = .063, 90% CI, .000–.130; Comparative Fit Indices (CFI) = .998; Tucker-Lewis Index (TLI) = .991).

Conclusions: EHF-4 shows high internal consistency and a single dimension that explains more than 50% of the total variance. Further studies are needed to confirm these observations, that can be taken as preliminary.

© 2016 Asociación Colombiana de Psiquiatría. Published by Elsevier España, S.L.U. All rights reserved.

* Corresponding author.
E-mail addresses: mpoarias@comportamientohumano.org, campoarias@gmail.com (A. Campo-Arias).

Escala breve para homofobia en estudiantes de Medicina de dos universidades colombianas: resultados de un proceso de refinamiento

RESUMEN

Palabras clave:
Homofobia
Análisis factorial
Confiabilidad y validez
Estudiantes de Medicina
Estudios de validación

Introducción: La evaluación de escalas de medición es un proceso continuo que exige revisiones y adaptaciones según las características de los participantes. La Escala para Homofobia de siete ítems (EHF-7) ha mostrado aceptable desempeño general en estudiantes de Medicina de dos universidades en Colombia. No obstante, el desempeño de algunos ítems fue deficiente y se podría eliminar algunos con el mejoramiento en el comportamiento de los ítems conservados.

Objetivo: Revisar el funcionamiento psicométrico y afinar el contenido de EHF-7 en estudiantes de Medicina de dos universidades colombianas.

Métodos: Participaron 667 estudiantes de 18-34 (media, $20,9 \pm 2,7$) años, de los que el 60,6% eran mujeres. Se calcularon el alfa de Cronbach (α) y el omega de McDonald (Ω) como indicadores de confiabilidad, y para afinar la escala, se llevaron a cabo análisis de factores exploratorios (AFE) y confirmatorios (AFC).

Resultados: EHF-7 mostró $\alpha = 0,793$ y $\Omega = ,796$, además de un factor principal que explicó el 45,2% de la varianza total. Los AFE y AFC indican eliminar tres ítems. La versión de cuatro ítems (EHF-4) alcanzó $\alpha = 0,770$ y $\Omega = 0,775$, con un único factor que dio cuenta del 59,7% de la varianza total, con AFC que mostró mejores indicadores ($\chi^2 = 3,622$; gl = 1; p = 0,057; RMSEA = 0,063; IC90%, 0,000-0,130; CFI = 0,998; TLI = 0,991).

Conclusiones: EHF-4 presenta alta consistencia interna y una dimensión que explica más del 50% de la varianza total. Es preciso realizar investigaciones que confirmen estas observaciones aún preliminares.

© 2016 Asociación Colombiana de Psiquiatría. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

In Colombia, with the increasing visibility of non-heterosexual groups and the impact on people's physical and mental health when they perceive themselves to be stigmatised and discriminated against, the need for knowledge about the attitudes of other groups towards these people makes it necessary to have valid and reliable measuring instruments that quantify different types of prejudices.[1,2]

The Homophobia Scale (HS-7) is a seven-item, Likert-type questionnaire for quantifying one prejudice, i.e. a negative attitude towards homosexual people.[3] The HS-7 is one of the instruments available for quantifying people's attitudes towards homosexuality, and one of its main attributes is the small number of items[3] Despite its small size, it shows good psychometric performance, with high internal consistency and adequate types of validity.[4,5] The fact that the HS-7 is so quick to complete may explain why it is used so frequently in research involving higher education students around the world.[6–8]

The performance of the HS-7 in medical students in Colombia has been presented in three previous articles. In the first study published, 199 medical students from the first to fifth semesters of a university in Bogotá, Colombia, participated and the scale was reported to have a high internal consistency ($\alpha = 0.78$ and $\Omega = 0.79$), adequate convergent validity ($r = 0.84$ with the scale for attitude towards lesbians and gay men [ATLG]), acceptable discriminant validity ($r = -0.06$ with the

General Well-Being Index [WHO-5]), poor nomological validity ($r = 0.19$ with the short Francis scale for religiosity [Francis-5]) and one single domain or factor accounting for 44.7% of the variance.[9]

In the second study, 124 students participated, in this case from the sixth to the tenth semester at the same university in Bogotá, and the investigators found adequate internal consistency ($\alpha = 0.81$ and $\Omega = 0.82$), high convergent validity ($r = 0.82$ with ATLG), optimal discriminant validity ($r = -0.03$ with WHO-5), poor nomological validity ($r = 0.19$ with Francis-5; with no significant differences in the scores between men and women, when higher scores were expected from the men) and one single factor was retained that accounted for 49.2% of the total variance observed.[10]

Lastly, in the third study, 366 students from the first to the ninth semesters of a university in Bucaramanga participated; the findings were acceptable internal consistency ($\alpha = 0.78$ and $\Omega = 0.79$), good convergent validity ($r = 0.82$ with ATLG), optimal discriminant validity ($r = 0.03$ with WHO-5), inconsistent nomological validity ($r = 0.16$ with Francis-5, lower than expected, and with significant differences between men and women, higher amongst males, as is usual with most prejudices) and one single domain that accounted for 43.8% of the variance was retained. Also in that study, an additional validation test was performed, based on item-response theory: differential item functioning (DIF) was reported by gender, and no significant differences were found between males and females in any of the seven items of the scale.[11]

Factor analysis is usually related to the construct validity of a scale.[12] However, it should be borne in mind that all known and calculated validity forms contribute to the construct validity, the practical and objective utility of a theoretical concept.[12,13] The investigations previously reviewed showed that the factor solution was close to the desired 50% in only one of the analyses. It was also found that items 2, 4 and 6 showed poor individual performance, with corrected Pearson's correlations and low commonalities.[9–11]

Given that factor analysis is the most appropriate strategy for the review and refinement of measurement scales, an analysis was performed in this study to observe the performance of the HS-7 and to refine the scale after removing items with poor performance in the previous studies.[12] It was assumed that the validation of measuring instruments is a continuous process that requires constant revision and adaptation with the adequate use of different statistical tests.[12,13] In addition, in this secondary analysis the CFA (which was omitted in the preceding articles) was carried out to support the interpretation of the findings.

The objective of this analysis is to review the psychometric functioning and to refine the content of the HS-7 in medical students at two universities in Colombia.

Material and methods

We conducted an observational, analytical validation study within the context of a larger study that explored the psychometric performance of various scales in medical students. For this study, the standards for health research in Colombia were followed; an institutional ethics committee reviewed and approved the research project and the consent of the research participants was obtained once they had been informed about the objectives and the respect for privacy and confidentiality of the data provided.[14]

A total of 667 first to tenth semester medical students from two universities, one in Bogotá and another in Bucaramanga, participated in this study. The participating population was aged from 18 to 34 (mean, 20.9 ± 2.7) years. With regard to gender, 60.6% were women. This group represents 96.8% of the samples participating in the three studies mentioned above, as we excluded 22 (3.2%) participants who did not complete the Zung Self-Rating Anxiety Scale–Short Form (ZSAS-SF), which was included in the present analysis and showed better internal consistency than WHO-5, used in the other analyses for discriminant validity.[9–11] Participants completed the questionnaire in the classroom in the presence of a research assistant, who presented the study objectives, requested voluntary participation and gave instructions on how to complete the research questionnaire. The questionnaire did not ask for the person's name, with the aim being that completing it anonymously would encourage them to answer as honestly as possible. It asked only for basic demographic information and included the ZSAS-SF,[15] the scale for attitude towards gay men (ATG),[16] and the HS-7.[3]

The ZSAS-SF is a brief self-administered questionnaire that consists of five questions that investigate symptoms such as nervousness, fear for no apparent reason, muscular pains, easy fatigue and feeling dizzy in the period covering the last 30 days. The scale provides four response options ranging from never to always. The response of the participant is rated 1–4, with a range of possible total scores of 0–20; a higher score indicates more and greater anxiety symptoms that may be of clinical importance. This scale has been used in different research projects in Colombia and shows high internal consistency.[15]

The ATG is a ten-point scale that explores attitudes towards homosexual men in relation to different topics such as adoption, marriage, work and other general impressions. The instrument consists of a Likert-type response pattern (polytomous) with five response options from "strongly disagree" to "strongly agree". Each response is given a score of 0–4, a possible spectrum of 0–40. The higher the score, the worse the attitude towards gay men, with more extreme prejudice or homophobia.[16] The Spanish version of the scale shows good psychometric performance.[9,17,18]

Cronbach's alpha (α)[19] and McDonald's omega (Ω)[20] were calculated as reliability indicators. To determine the convergent and discriminant validity, Pearson's correlation (r) was calculated.[21] For convergent validity, the total scores on the ATG and HS-4 were correlated, and for the discriminant validity, the total scores on the ZSAS-SF and HS-4.

For the estimation or verification of the nomological validity of the HS-4, the means $\pm$ standard deviation of the male and female scores were compared using the Student's t test (significantly higher scores were expected from men than from women).

Lastly, in order to corroborate the dimensionality of HS-4 and HS-7, a factor analysis was carried out by the maximum likelihood method, the commonalities were observed and the Kaiser–Meyer–Olkin (KMO) coefficient[22] and Bartlett's test of sphericity of the sample were calculated.[23] For the KMO coefficient, a value >0.600 was expected and for the Bartlett's test, probability <0.0524.

CFA was performed to confirm the factor structure previously determined in the EFA. In order to evaluate the fit of the models in HS-7 and HS-4, the χ^2 test was determined with the respective degrees of freedom (df) and probability value (p), root mean square error of approximation (RMSEA), with a 90% confidence interval (90% CI) as is customary, the comparative fit index (CFI) and the Tucker-Lewis index (TLI). For χ^2 the probability value was expected to be >5%; for RMSEA, <0.06, and for CFI and TLI, values >0.89. Most of this analysis was performed with the SPSS 16.0 statistical package,[25] while the CFA was completed with the Mplus 7.21 software.[26]

Results

The HS-7 showed $\alpha = 0.793$ and $\Omega = 0.796$, and a main factor that explained 45.2% of the total variance. The CFA showed $\chi^2 = 139.756$; $df = 13$; $p < 0.01$; RMSEA = 0.121; 90% CI, 0.103–0.139; CFI = 0.953; TLI = 0.923.

Given that these findings indicated the removal of three items, the performance of a four-item version (HS-4) was tested. The HS-4 showed $\alpha = 0.770$ and $\Omega = 0.775$, with one single factor accounting for 59.7% of the total variance ($\chi^2 = 3.622$; $df = 1$; $p = 0.057$; RMSEA = 0.063; 90% CI, 0.000–0.130; CFI = 0.998;

Table 1 – Commonalities and coefficients of the items in the CFA.

Items	HS-7 (KMO = 0.827)		HS-4 (KMO = 0.768)	
	Commonality	Coefficient	Commonality	Coefficient
1. Uncomfortable	0.379	0.615	0.398	0.631
2. Honest	0.152	0.390		
3. Corrupt	0.424	0.651	0.436	0.660
4. Right	0.343	0.586		
5. Sin	0.458	0.676	0.507	0.709
6. Contribution	0.275	0.524		
7. Illegal	0.523	0.723	0.518	0.720
Actual value	3.164		2.388	
Total variance (%)	45.2		59.9	

Bartlett's: HS-7 = 1187.525; df = 21; $p < 0.001$; HS-4 = 679.448; df = 6; $p < 0.001$.

TLI = 0.991). The details on commonalities and coefficients of the items in the factor analysis are shown in Table 1.

For the nomological validity of HS-4, we compared the mean ± standard deviation of the male and female scores (10.1 ± 3.7 vs. 9.3 ± 3.6). The difference was statistically significant (Levene's test for homogeneity of variance, $F = 0.004$, $p = 0.949$, $t = 2.499$, $df = 665$, $p = 0.013$, two-tailed).

The convergent validity of HS-4 with ATG ($\alpha = 0.821$) showed high correlation ($r = 0.778$, $p < 0.001$). The discriminant validity with ZSAS-SF ($\alpha = 0.789$) showed a very poor correlation ($r = -0.047$; $p = 0.223$).

Discussion

This research shows that HS-7 and HS-4 are scales with high reliability and adequate validity. However, the shorter version accounts for a higher percentage of the total variance, with better indicators in the CFA.

Instruments need to be validated to improve the measurements of constructs, both in the clinical context and in research work.[12,27,28] The tendency at the moment is to continually revise the scales already available, with careful evaluation of the performance of individual items and removal of those with poor indicators, in order to reduce the number of items on the scales without losing reliability, validity and practical utility.[12,13,27,29,30]

There are a number of advantages to short-form instruments in clinical and epidemiological studies. First, from the psychometric perspective, is that they collect or preserve the essential or structural aspects of the construct that usually meet in the factor or dimension that explains the main or highest percentage of the total variance.[24,31] Second, use of these shorter versions reduces the possibility of overestimating the reliability and internal consistency due to the number of points in the scale, as that type of coefficient is sensitive to the number of items. The greater the number of items, the greater the internal consistency, even with a significant reduction in the intercorrelations among items, which to a large extent is an indicator that the different points approach or attempt to quantify the same construct[32,33] Third and last is the fact that it is more operative. A measuring instrument should be practical both for application and for the qualification and interpretation of the scores. Short scales reduce the

amount of time required for completion, with less chance of users developing fatigue or boredom, as can occur with long questionnaires, thus providing even more assurance of the validity and reliability of the measurement.[27]

Having instruments such as the HS-4 with good psychometric performance in medical students is necessary, given the high frequency of sexual prejudice in this group of people.[34] It will allow research to be carried out to help identify the scale of the problem, so that the necessary appropriate measures can be taken to reduce the negative impact of sexual prejudices in the medical profession during the training process and while practising.[35]

We are able to conclude that in medical students from two cities in Colombia, the HS-4 showed high internal consistency, good convergent validity, adequate discriminant validity, excellent nomological validity and one dimension that explains more than 50% of the total variance, with better indicators in the CFA fit than the HS-7. Further research is needed to show the psychometric performance of the IHS-4 and confirm these initial observations, which must be considered as preliminary.

Ethical disclosures

Protection of human and animal subjects. The authors declare that the procedures followed were in accordance with the regulations of the relevant clinical research ethics committee and with those of the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Confidentiality of data. The authors declare that they have followed the protocols of their work centre on the publication of patient data.

Right to privacy and informed consent. The authors have obtained the written informed consent of the patients or subjects mentioned in the article. The corresponding author is in possession of this document.

Conflict of interests

The authors declare that they have no conflicts of interest.

Acknowledgements

REFERENCES

1. Paradies Y. A systematic review of empirical research on self-reported racism and health. Int J Epidemiol. 2006;35:888–901.
2. Alvarez-Galvez J, Salvador-Carulla L. Perceived discrimination and self-rated health in Europe: evidence from the European Social Survey (2010). PLOS ONE. 2013;8:e74252.
3. Bouton RA, Gallager PE, Garlinghouse PA, Leal T, Rosenstein LD, Young RK. Scales for measuring fear of AIDS and homophobia. J Pers Assess. 1987;51:606–11.
4. Grey GA, Robinson B, Coleman E, Bockting WO. A systematic review of instruments that measure attitudes toward homosexual men. J Sex Res. 2013;50:329–52.
5. Costa AB, Bandeira DR, Nardi HC. Systematic review of instruments measuring homophobia and related constructs. J Appl Soc Psychol. 2013;43:1324–32.
6. Miller DB, Briggs H, Corcoran K. Fear of AIDS and homophobia scales: additional estimates of reliability and validity. Psychol Rep. 1997;81:783–6.
7. Aberson CL, Swan DY, Emerson EP. Cover discrimination against gay men by U.S. college students. J Soc Psychol. 1999;139:323–34.
8. Long W, Millsap CA. Fear of AIDS and Homophobia Scale in an ethnic population of university students. J Soc Psychol. 2008;148:637–40.
9. Campo-Arias A, Lafaurie MM, Gaitán-Duarte HG. Confiabilidad y validez de la escala para homofobia en estudiantes de medicina. Rev Colomb Psiquiatr. 2012;41:867–80.
10. Campo-Arias A, Oviedo HC, Herazo E. Escala para homofobia: Validez y confiabilidad en estudiantes de medicina de una universidad de Bogotá (Colombia), 2010. Arch Med. 2014;14:9–20.
11. Campo-Arias A, Oviedo HC, Herazo E. Desempeño psicométrico de la escala para homofobia en estudiantes de medicina. Rev Cienc Biomed. 2014;5:287–94.
12. Reise SP, Waller NG, Comrey AL. Factor analysis and scale revision. Psychol Assess. 2000;12:287–97.
13. Strauss E, Spreen O, Hunter O. Implications of test revisions for research. Psychol Assess. 2000;12:237–44.
14. Ministerio de Salud. Resolución 8430 por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Bogotá: Ministerio de Salud; 1993.
15. De la Ossa S, Martínez Y, Herazo E, Campo A. Estudio de la consistencia interna y de la estructura factorial de tres versiones de la escala de Zung para ansiedad. Colomb Med. 2009;40:71–7.
16. Herek GM. Heterosexuals' attitudes toward lesbians and gay men: correlates and gender differences. J Sex Res. 1988;25:451–77.
17. Cárdenas M, Barrientos JE. The attitudes toward lesbians and gay men scale (ATLG): adaptation and testing the reliability and validity in Chile. J Sex Res. 2008;45:140–9.
18. Barrientos JE, Cárdenas M. A confirmatory factor analysis of the Spanish language version of the Attitudes Toward Lesbians and Gay Men (ATLG) Measure. Univ Psychol. 2012;11:579–86.
19. Cronbach J. Coefficient alpha and the internal structure of test. Psychometrika. 1951;16:297–334.
20. McDonald RP. Theoretical foundations of principal factor analysis and alpha factor analysis. Br J Math Stat Psychol. 1970;23:1–21.
21. Pearson K. Determination of the coefficient of correlation. Science. 1909;30:23–5.
22. Kaiser HF. An index of factorial simplicity. Psychometrika. 1974;34:31–6.
23. Bartlett MS. Test of significance in factor analysis. Br J Psychol. 1950;3:77–85.
24. Campo-Arias A, Herazo E, Oviedo HC. Análisis de factores: fundamentos para la evaluación de instrumentos de medición en salud mental. Rev Colomb Psiquiatr. 2012;41:659–71.
25. SPSS for windows 16.0. Chicago: SPSS. Inc; 2006.
26. Muthén LK, Muthén BO. Mplus for Windows 7. 31. Los Angeles: Muthén & Muthén; 2015.
27. Sánchez R, Echeverry J. Validación de escalas de medición en salud. Rev Salud Publica. 2004;6:302–18.
28. Rodríguez MA, Lopera J. Conceptos básicos en la validación de escalas en salud mental. Rev CES Med. 2002;16:31–9.
29. San Martín S. Depuración de escalas en marketing. Metodol Encuestas. 2003;5:183–209.
30. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. J Nurs Sch. 2007;39:155–64.
31. Norman GR, Streiner DL. Bioestadística. Madrid: Mosby/Doyma libros; 1996.
32. Roberts P, Priest H, Traynor M. Reliability and validity in research. Nurs Stand. 2006;20:41–5.
33. Campo-Arias A, Oviedo HC. Propiedades psicométricas de una escala: la consistencia interna. Rev Salud Publica. 2008;10:831–9.
34. Campo-Arias A, Herazo E. Homofobia en estudiantes de medicina: una revisión de los diez últimos años. Medunab. 2008;11:120–3.
35. Sanchez NF, Rabatin J, Sanchez JP, Hubbard S, Kalet A. Medical students' ability to care for lesbian, gay, bisexual, and transgendered patients. Fam Med. 2006;38:21–7.