Structural validation and multiple group assessment of the short Internalized Homonegativity scale in homosexual and bisexual men in 38 European countries. Results from the European MSM Internet Survey

Ha Tran[1], Michael W. Ross[2], Pamela M. Diamond[1], Rigmor C. Berg[3], Peter Weatherburn[4], and Axel J. Schmidt[4].

[1]Center for Health Promotion and Prevention Research, School of Public Health, University of Texas, PO Box 20036, Houston, Texas 77225, USA.

[2]Program in Human Sexuality, Department of Family Medicine and Community Health, University of Minnesota Medical School, 1300 South 2nd Street, Minneapolis, MN 55454, USA and Division of Social Medicine and Global Health, Lund University, Clinical Sciences, Malmö, Sweden.

[3]Department of Public Health, University of Tromsø, and the Norwegian Knowledge Center for Health Services, Nasjonalt Kunnskapssenter for Helsetjenesten, Boks 7004, St. Olavsplass 0130 Oslo, Norway.

[4]Sigma Research, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, U.K.

Abstract

Internalized Homonegativity (IH) is the internalization of negative attitudes and assumptions about homosexual people by homosexual people themselves. To measure IH, Smolenski, Diamond, Ross, and Rosser (2010), and Ross, Rosser, and Smolenski (2010), revised *The Reactions to Homosexuality Scale* (RHS) to develop the *Short Internalized Homonegativity Scale* (SIHS) with 8 items. Using the European Men-who-have-sex-with-men Internet Survey (EMIS) data with an analytic sample of 130,718 gay and bisexual men in 38 European countries, we confirmed the validity of SIHS scale in both training and validation data, in strata of Ross et al. (2013)'s three "homosexual discrimination" country clusters, of age, and of education level. However, the performance was less adequate in comparison of gay versus bisexually-identified individuals. The latent SIHS structure contained only minor variations across these three strata. The 7-item scale performed as well as the 8-item scale. SIHS is a promising candidate for standard IH measures, which is invariant across cultural, age, and educational strata.

Key terms: Internalized Homonegativity, Internalized Homophobia, measurement invariance, EMIS, MSM, homosexual.

Introduction

Internalized Homonegativity (IH: previously known also as Internalized Homophobia) is defined by Ross and colleagues as the internalization of negative attitudes and assumptions about homosexual people by homosexual people themselves (Ross, Berg, et al., 2013). More general definitions such as negative attitudes and assumptions about sexual minority individuals by sexual minorities themselves may be used, but it is unclear whether there is sufficient common ground in measurement dimensions between IH in gay men and lesbians, or trans individuals, or even between gay/homosexual and bisexual men, to allow for broader definitions to be psychometrically useful. Previous research has shown the relationships of IH with distrust and loneliness (Shidlo, 1994; Weber, 2005), the use of illegal substance and alcohol consumption (Finnegan & Cook, 1984; Meyer & Dean, 1998; O'Halleran Glaus, 1988), defense mechanisms such as rationalization, denial, and identification with the aggressor (Currie, Cunningham, & Findlay, 2004; Margolies, Becker, & Jackson-Brewer, 1987), avoidance of HIV testing (Shoptaw et al., 2009), decrease of condom use (Huebner, Davis, Nemeroff, & Aiken, 2002), and commercial sex (Ross, Kajubi, Mandel, McFarland, & Raymond, 2013). As Ross and Rosser (1996) and Currie et al. (2004) affirmed, IH has a central role in working with health related risk factors of non-heterosexual populations, especially men (Currie et al., 2004; Ross & Rosser, 1996). Given increasing cross-cultural research on IH, and the rise in use of Internet platforms which often require shorter measures to retain participants, it is important to establish the stability of short IH measures across cultural and demographic parameters in gay and bisexual men. This paper reports on a study investigating the structure and structural invariance by age, education, state of residence, legal climate, and gay versus bisexual orientation, of a short IH measure.

In this paper, we use the term "homosexual" as a gender-neutral adjective to describe identity or behavior, or discrimination against these, but in some of the 38 countries we surveyed in Europe, many men describe their identity as homosexual men (the option in the questionnaire was "gay or homosexual"). The term "gay" may be taken to refer to the western lifestyle of the same name but not anything that would resemble their existence or identity. "Homosexual" man and its translations were terms also used in the EMIS questionnaire ("gay" in many languages does not translate), and rather than impose what may be considered an alien or Americo-colonial term, not necessarily synonymous with homosexual, and characterized by an English-language loan-word, we used a term that was understood and spoken by many of the men themselves. In this paper we also use the term LGBQ+ (Lesbian, Gay, Bisexual and "Queer" and related definitional terms including homosexual) to describe sexual orientation, and the term MSM to describe homosexual behavior.

Multiple investigators have attempted to develop IH scales to capture the dimensions of Internalized Homonegativity (Bell & Weinberg, 1978; Mayfield, 2001; Nungesser, 1983; Ross & Rosser, 1996; Shidlo, 1994). The Nungesser Homosexual Attitudes Inventory (NHAI; Nungesser, 1983) was the most widely used scale to measure IH. This 34-item instrument consists of three subscales: Self, Other, and Disclosure. The psychometric properties of NHAI were evaluated as good with a coefficient alpha of 0.94 for the entire scale (Mayfield, 2001). However, as this scale was constructed 30 years ago, its validity has been questioned given large societal changes and the increasing acceptance of LGBQ+ individuals in the Western world. In addition, some of items were measures of antecedents or consequences of IH, such as the items "I do not think I will be able to have a long-term relationship with another man" and "If others knew of my homosexuality, I would not be afraid that they would see me as being effeminate" (Mayfield, 2001). Mayfield (2001) developed the Internalized Homonegativity

Inventory (IHNI) instrument which consisted of 23 items comprising three factors, namely, Personal Homonegativity, Gay Affirmation, and Morality of Homosexuality. Although the scale obtained a good internal consistency with a Cronbach's alpha of 0.91 for the entire scale, some items such as "Sometimes I feel that I might be better off dead than gay" appeared extreme and overt in their assessment (Currie et al., 2004). Extreme items were possibly the reason for the skewed distribution of scores reported in the original study (Currie et al., 2004).

Among the IH scales, the Reaction to Homosexuality Scale (RHS) with 26 items loading on four factors ("public identification as gay," "perceptions of stigma associated with being gay," "social comfort with gay men," and "moral and religious acceptability of being gay") was considered as a more sensitive measure of IH (Currie et al., 2004; Ross & Rosser, 1996; Williamson, 2000). However, several items of RHS only measured the constructs conceptually relating to IH. For example, the items "I worry about becoming unattractive," "Discrimination against gay people is still common," and "Most of my friends are homosexual" did not directly address IH. In addition, the scale was relatively long, making it difficult to include in surveys and possibly leading to missing data (Currie et al., 2004; Smolenski et al., 2010).

In 2004, Currie and coworkers (Currie et al., 2004) revised the RHS to overcome some of its limitations. This revised measure consisted of 12 items with three factors: "public identification as gay," "sexual comfort with gay men," and "social comfort with gay men" (Currie et al., 2004). Although this scale greatly decreased the number of items, the same problems were present as in the original RHS scale, that some items only indirectly measured IH through conceptually related factors (Smolenski et al., 2010). In addition, Currie et al. (2004) did not analyze the invariance of IH structure

across population strata, which left the questions about the external validity of this measure open. These issues inhibited the application of this revised measurement in IH research.

In 2010, another revised version of the RHS developed by Smolenski et al., overcame these two problems (Figure 1). This revised IH measure consisted of eight items loading on three factors: "social comfort with gay men," "public identification as gay," and "personal comfort with a gay identity" (Smolenski, 2009). This short form of the IH scale (SIHS) showed a good reliability score and good fit indices in both training and validation datasets. This latent IH structure was also validated to be consistent in the goodness of fit across population strata of race/ethnicity and languages in a U.S. self-identified Latino MSM population. Its validity also was also confirmed in a Ugandan MSM population (Ross, Smolenski, et al., 2010). In these psychometric analyses, Smolenski (2009) and Smolenski et al. (2010) argued that there was limited psychometric benefit of including one item, "Obviously effeminate homosexual men make me feel uncomfortable." Comparing comparability of the English and Spanish versions, Smolenski et al. (2010) removed this item for having a cross-loading with the Personal Comfort with a Gay Identity factor, which suggested low discriminatory ability of the item, reducing the 8-item scale to 7 items.

With an urgent need of IH research in developing a standard IH scale that is relevant internationally, SIHS should be further confirmed in MSM populations of multiple countries. SIHS was used in the European MSM Internet Survey (EMIS), the largest MSM survey to date. EMIS covered 38 countries, was disseminated in 25 languages, and included 174,209 participants (Weatherburn et al., 2013; The EMIS Network, n.d.). This study thus provides an opportunity to confirm the SIHS in a very large and diverse sample of MSM.

In addition to validating the SIHS in pan-European data, it was necessary to explore whether this measure captured the same latent variables across population strata. Without achieving the minimum requirements of measurement invariance, the IH study needed to perform under specific circumstances of IH structure for each subsample (Stain, Lee, & Jones, 2006). Given the wide differences in social norms and prejudice against gay/homosexual people in European countries, researchers wanted to explore the measurement non-invariance across "homosexual discrimination" country clusters which were derived using clustering analysis (Berg, Ross, Weatherburn, & Schmidt, 2013; Berg, Weatherburn, Ross, & Schmidt, 2015; Ross, Berg, et al., 2013). This will determine if IH is dimensionally similar in high, medium and low homonegative cultures and thus whether it is valid to compare IH scores between settings which range from liberal to reactionary in their response to gay/homosexual people.

Assessment of measurement invariance of age strata was also the researchers' interest. Ross and colleagues reported an association between IH and age in EMIS data (Ross, Berg, et al., 2013). Lingiardi, Biocco, and Nardelli (2012) also found that IH decreased significantly with age. The exploration of differences of IH measure in age strata was considered one important demographic dimension to explore as gay men probably increasingly accepted their sexual orientation after time.

We additionally investigated measurement non-invariance as a function of education due to cultural attitudes about gay men. Higher educated homosexual men were more likely to live in environments and communities in which LGBQ+ people were less stigmatized. They themselves probably had less internalized homophobia due to more exposure to information on homosexuality and supportive social and legal contexts.

Because of the need to confirm the appropriateness of SIHS in an international MSM sample such as EMIS, and the necessity to evaluate the measurement equality by "homosexual discrimination"

country groups, and by age and education strata, we proposed a study with three aims: (1) confirm the structure of the SIHS developed by Smolenski, 2009; (2) assess the measurement invariance by country clusters defined by level of homosexual discrimination (Ross, Berg, et al., 2013), age, and education strata; and (3) compare the 8-item version with the 7-item version proposed by Smolenski (2009) and Smolenski et al. (2010), extending that comparison from the Spanish-English comparison of Smolenski (2010) to the other demographic variables of age, education, and liberal-moderate-conservative state residence. The overall goal of this study was to demonstrate the validity of SIHS scale in a very large MSM data set.

Methods

The European MSM Internet Survey (EMIS) was carried out by a large network of European organizations, led by Robert-Koch-Institute (Germany) and Sigma Research (now London School of Hygiene and Tropical Medicine), with institutional review board approval through the University of Portsmouth, England. From June through August 2010, the study was promoted via non-governmental organizations in each participating country, and through invitations in gay social media and a wide variety of over 235 (trans-)national websites for MSM. Five large dating sites sent instant messages: PlanetRomeo®, Manhunt/Manhunt Cares®, and Gaydar®, each of which has membership across Europe; Qguys® for (Russian-speaking) countries within the Commonwealth of Independent States and the Baltic countries; and Qruiser® for Scandinavia. A core slogan was used to promote the survey, the English language version of which was "Be part of something huge!" The study's welcome webpage gave potential participants an option of 25 languages. Upon selection of language, the study website described the research in the chosen language, and eligible volunteers were routed to the survey questions.

Eligible participants were men who were legally of age to have consensual sex with men in their country of residence and who were attracted to men and/or had sex with men. This included men who self-identified as homosexual, gay, bisexual, or heterosexual. Additionally, the men needed to live in Europe and declare that they had read and understood the aim of the study. Respondents were not compensated for their participation; no IP address was recorded. The EMIS methods are described in detail elsewhere (Weatherburn et al., 2013, The EMIS Network, n.d.).

The sample selection for this analysis is based on 3 exclusion criteria: (1) Due to a translation error, the Spanish language questionnaire was missing the first IH item "Obviously effeminate homosexual men make me feel uncomfortable" (I1, Table 2); therefore all respondents who answered the questionnaire in Spanish were excluded (*n*=12,934). Answers from Spain are thus based solely on respondents living in Spain but using a different language than Spanish. (2) Respondents with missing answers to this first item (*n*=1,710) or any of the other 7 IH items (*n*=27,980) were also excluded. In contrast to previous implementations of the SIHS, respondents' missing answers include the additional answer option "does not apply to me," which was added because the EMIS network thought it would be odd to ask men if they felt comfortable in gay bars when living in a country where gay bars do not exist, based on the EMIS piloting phase. The proportion of missing answers ranged from 12.6% in English to 38.1% in Polish; and from 12.5% in West Europe (http://www.emis-project.eu/sub-regions.html) to 27.8% in the Eastern part of Central Europe (The EMIS Network, n.d., pp.169-170). (3) Trans respondents (*n*=393) as well as men with no evidence of sexual attraction to men (*n*=465) were also excluded. The final analytic sample was 130,718. Selected demographics are provided in Table 1.

Measures

Internalized Homonegativity – All eight items (Table 2) were measured by a 7-point (scored 0-6) Likert-type scale from "strongly disagree" to "strongly agree." Item 1 and item 3 were reverse-coded. All items were scaled to show an increase in IH with an increased manifest score.

"Homosexual Discrimination" country clusters – 38 countries were divided into three groups: 12 "conservative" countries, 17 "moderate" countries, and 9 "liberal" countries by clustering analysis. This clustering analysis was performed and reported by Ross, Berg, et al., 2013, based on three variables: the proportion of EMIS respondents experiencing verbal discrimination; the proportion of EMIS respondents reporting physical violence because of assumed homosexual behavior or identity; and a measure of legal climate supporting LGBQ+ rights. The latter was measured by scoring the presence of six legislative measures of LGBQ+ status including legality of homosexual acts, recognition of same-sex relationships, same-sex marriage, same-sex adoption, gays can serve in the military, and LGBQ+ antidiscrimination laws. Legal climate was measured by scoring the presence of the six legislative measures of LGBQ+ status (legal discrimination) listed by Wikipedia, with a high score of six for the presence of all legislative protections (homosexual acts are legal, recognition of same-sex relationships, same-sex marriage, same-sex adoption, gays serve in the military, LGBQ+ antidiscrimination laws).

These three country clusters were then used to test whether SIHS had the same measurement characteristics across the three (liberal, moderate, conservative) politically disparate country clusters. The list of countries in each group is displayed in Table 1.

Education was measured by using the 1997 version of the International Standardized Classification of Educational Degrees (ISCED) levels. The median ISCED level was 5 (first stage of tertiary education), which was used to divide the analytic sample into two groups (with or without tertiary education) based on education level in multi-group invariance analysis.

Analyses

The analyses were conducted by using Mplus 7.11 (Muthén & Muthén, 2013) and SPSS V.19 (IBM, Chicago, Illinois, USA) on the 38-country dataset with 130,718 observations. We randomly divided the sample into halves: a training sample and a validation sample. Confirmatory factor analysis (CFA) was used for training dataset to confirm the validity of the SIHS developed by Smolenski (2009). Maximum Likelihood with the robust standard errors and chi square (MLM) estimators was used because of the non-normality of the eight scale items.

We assessed the goodness of fit by using fit indices Comparative Fit Index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Bentler & Bonett, 1980), and the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990). We considered that a model had a good fit or acceptable fit where CFI and TLI are larger than 0.95 or 0.90, respectively, and RMSEA is less than 0.08 (good fit) or 0.05 (acceptable fit) (Hu & Bentler, 1999). Because the chi-square goodness of fit tests are prone to a high type 1 error rate when sample size is large, these tests were not considered in fit assessment of measurement models (Milfont & Fischer, 2010).

We estimated a series of nested multi-group models to test for measurement and structural invariance across subgroups of respondents defined by "Homosexual Discrimination" country clusters, age, and education. The procedure of the measurement invariance tests has been described in several reports elsewhere (Brown, 2006; Byrne, 2012; Byrne & Stewart, 2006; Stain et al., 2006). In brief, the first level of invariance, configural invariance, required only the same number of factors and the same loading patterns across groups. The second level of invariance, metric invariance, constrained the equality of factor loadings in all groups. To assess the third level, scalar invariance, the item intercepts were constrained to the equality of intercepts across all groups. We also tested strict invariance by

constraining the invariance of residuals and residual covariance. Lastly, we assessed the invariance of second order factor loadings following the recommendation of Byrne & Stewart (2006). This latter approach is important because measurement equivalence is of particular concern in cross-cultural research whereby assessment may be based on a translated version of the original instrument. For this reason, we did not want to prejudge whether differences were due to true attitudinal differences or, rather, to psychometric differences related to the item responses (Cheung & Rensvold, 2002). We repeated all analyses for the 7-item version of the SIHS suggested by Smolenski et al. (2010).

Due to the sensitivity of the large sample size to the likelihood ratio test, we did not use this test to identify whether the more restricted model has worse fit than the less restricted model. We applied the recommendations of Cheung & Rensvold (2002) on the change of CFI and RMSEA estimates (less than or equal to -0.01 and larger than 0.15 indicating a non-substantial change in fit, respectively) between nested and comparison models.

Results

Table 1 depicts the demographic characteristics of the EMIS sample, as well as the mean IH score per country. Using a convenience sampling method, the analytic sample data from 38 countries was different across promoted countries in sample size (ranging from 86 participants in Moldova to 44,339 participants in Germany), in median age (ranging from 24 years in Moldova to 40 years in the Netherlands), and in percent with tertiary education (ranging from 31.1% in Austria to 85.6% in Turkey). The mean IH score also varied from 1.21 in the Netherlands to 2.57 in Bulgaria.

Table 2 presents the means, standard deviations, standardized coefficients, and standard errors of standardized coefficients of the eight scale items. As noted earlier, item 1 (Obviously effeminate homosexual men make me feel uncomfortable) and item 3 (Social situations with gay men make me feel

uncomfortable) were reverse coded to be on a consistent scale with the other items. The means of the eight items range from 1.50 (item 7) to 4.29 (item 1). These eight items loaded into three factors "Social comfort with gay men" (SC), "Public identification as gay" (PUBID), and "Personal comfort with a gay identity" (PC). The indicators loaded strongly on the common factors, with the lowest pattern coefficient 0.36 (item 1 on SC), and the highest pattern coefficient 0.84 (item 7 and item 8 on PC). The standardized loading factors of three common factors SC, PUBID, and PC on the second- factor variable IH are also high (0.75, 0.89, and 0.91, respectively). The correlations between items were from low to moderate.

Figure 2 describes the confirmatory factor model estimated using the training dataset. This model had a good fit, with CFI=0.96, TLI=0.93, and RMSEA=0.06. This model had a coefficient of internal consistency (Cronbach's alpha) of 0.77, which indicated acceptable reliability of this scale. This factor model also had an acceptably good fit in the validation sample (CFI=0.96, TLI=0.93, and RMSEA=0.06). Cross-validation of this final model was confirmed by testing for measurement invariance across the training and validation samples (Table 3). All of measurement invariance tests including configural invariance, metric invariance, scalar invariance, residual invariance, and second factor loading invariance between the training sample and validation sample were satisfied when the ΔCFI and ΔRMSEA were under the cutoff points (≥-0.01 for ΔCFI and ≤0.15 for ΔRMSEA). As we employed the estimator MLM, not ML, for factor analysis, the chi-square value of the configural model was obtained as the approximation, instead of the accuracy, of the sum of the subgroups' chi-square values. The final factor model was also tested for invariance across the three country clusters (Table 4). The factor model had a reasonable fit in both "liberal" country clusters" (CFI=0.97, TLI=0.94, and RMSEA=0.06) and "moderate" country cluster (CFI=0.96, TLI=0.93, and RMSEA=0.06). However,

this model had a somewhat poorer fit in the "conservative" country cluster, with CFI=0.94, TLI=0.88, and RMSEA=0.07. In the subsample "conservative" country cluster, adding the correlation between I1 and I3 greatly improves the goodness of fit of this model (CFI=0.99, TLI=0.98, and RMSEA=0.03). Once acceptable models were developed for each of the country clusters, measurement invariance was assessed. The models imposing configural invariance and metric invariance showed acceptable fit (the model numbers reported in Table 4). However, scalar invariance was not confirmed. According to the CFI difference, the model in which all items' intercepts were constrained to equality across groups showed worse fit compared with the less restricted metric invariance model (ΔCFI=-0.03). Freeing the intercepts to vary across groups for items 1, 2, and 3 resulted in acceptable fit for the model based on absolute fit indices as well as difference in indices. Constraining all residuals except the one for item 3 resulted in acceptable fit across country clusters. As the CFA model had a second ordered structure, we also constrained the second order factor loadings which were shown to be invariant across subpopulations. Using the 7-item model did not alter the finding, with the exception of in the 7-item structure; it was found that factor PC (personal comfort with a Gay identity) is correlated with PUBID (public identification as gay).

The confirmatory factor model was also applied to the various age and education groups (Table 5). Except for the variance of the intercepts of item 4 and item 8 between two groups of age (less than 33 years old and greater than or equal to 33 years old), the measurement invariance tests result in adequate to good fit. This result showed that the intercepts of item 4 (younger population: 1.43; older population: 1.35) and item 8 (younger population: 1.24; older population: 1.22) vary between younger MSM and older MSM, indicating partial measurement invariance. Between the subgroups determined

by the education variable, the factor model was confirmed to be invariant between the lower educated MSM population and the higher educated MSM population.

Table 6 describes the confirmatory factor model's application to the gay versus bisexual samples. When the measurement invariance tests for both 8-item and 7-item structures were performed across two groups of gay and bisexual men, the intercepts of most items were non-invariant. Between gay and bisexual groups, there were differences in the levels of comfort in being at gay bars, being seen in public with an obviously gay person, and discussing homosexuality in a public situation. The levels of the two group's views that homosexuality was morally acceptable, and wishes to change their own sexual orientations, were also different.

Discussion

In this research, we confirmed the validity of the second order factor model developed by Smolenski et al. (2010), and Ross, Rosser, et al. (2010). The SIHS contained eight items which loaded on three first order factors "Social comfort with gay man" (SC), "Public identification as gay" (PUBID), and "Personal comfort with a gay identity" (PC). There was a second order general factor, suggesting a single higher-order dimension we have labelled "Internalized Homonegativity". This psychometric scale was used in the EMIS dataset, which contains data from 38 European countries.

Comparison of the 8-item scale and the 7-item scale indicated that the 7-item scale performed as well as the 8-item scale, confirming the removal of one item as suggested by Smolenski et al. (2010). We recommend using the 7-item scale and treating it as a continuous variable.

While we anticipated some differences in the factorial structure of IH across three "Homosexual Discrimination" country clusters, we only identified non-invariance in the intercepts of three items of the factor "Social comfort with gay men" and these differences were fairly small. On the basis of

difference in homosexual discrimination levels of country clusters, the inequalities in the baseline conditions of items "Obviously effeminate homosexual men make me feel uncomfortable," "I feel comfortable in gay bars," and "Social situations with gay men make me feel uncomfortable" were plausible. In some of the European countries, especially in Eastern ones, the gay bars might not even exist, or even publicly appearing as gay men might lead to arrest, possibly leading to the difference in gay men's comfort to expose their sexual interests. Regarding the strict inequality, as the residual invariance is considered too strict in most situations, we considered that the partial measurement invariance in the residuals of item 3 was very minor.

As noted previously, the only non-invariance between two populations of younger and older populations were two intercepts of item 4, "I feel comfortable being seen in public with an obviously gay person" and item 8 "Even if I could change my sexual orientation, I wouldn't." These differences are explainable as after a number of years, the gay men probably become more self-accepting about their sexuality as they met more gay men and received more social support. Therefore, older MSM might be more comfortable being seen in public and less desirous of changing their sexual attraction. In addition, while we expected that different gay men's education resulted in differences in internal and external discrimination against homosexual identity or behavior, we identified the complete measurement invariance in IH latent between lower educated and higher educated MSM.

Findings comparing the gay with the bisexual group, however, were less clear, and suggest that the SIHS may not adequately cover differences in sexual minority status between homosexually/gay and bisexually-identified men in terms of internalization of sexual minority status. Five of the seven items in the scale reflected non-invariance between the groups, and these data strongly suggest that internalization of a bisexual minority status is conceptually different from a gay man's minority status,

at least in terms of the items that might capture this, and their psychometric properties. Qualitative research with bisexual men to describe the internalized stigma associated with bisexuality is needed to better understand the nature of internalized biphobia.

As in the recommendations by Byrne, Shavelson, and Muthén (1989), full scalar invariance was not required for further tests of invariance and substantive analysis. They indicated that full scalar invariance was not necessary for further assessment of latent mean invariance. As in this study we did not aim to conduct tests to compare latent means across subsamples, the full invariance of intercepts (and residual invariance) was not required. Therefore, although some non-invariances in IH's CFA model appeared, either across country clusters or across population strata of age, we suggest that this scale could be used in pluralistic populations without introducing bias.

This research had some advantages. First, the self-completion internet-based survey was the most appropriate design for this study for several reasons: (1) this study design recruited a large sample of a "hidden" population of MSM over a large geographic area; (2) this study design was convenient for respondents as they were able to answer the questions on their own schedule and privately. The subjects could start the survey at one time, then stop and continue later; (3) this study design greatly reduced the cost and time for data collection (Weatherburn et al., 2013). EMIS recruited 174,209 subjects in little more than two months with a limited cost. Second, the study analytic applied the latent variable analysis with multi-group models which had the capacity to demonstrate the appropriateness of this scale in pluralistic populations.

Beside advantages, there were several limitations in this investigation. Some of disadvantages came from the survey design: (1) the study population was clearly restricted to those with access to a computer and the internet. This population tends to be younger, more educated, and urban (The EMIS

Network, n.d.), as has been previously shown by Ross, Månsson, Daneback, Cooper, & Tikkanen, 2005; (2) despite its size, the study population is not representative of MSM in Europe. Besides the study design, the study analytic sample included some disadvantages. First, a part of study population contained missing values in IH items and we had to rely on the available valid data. While the results were in agreement with our hypotheses and in previous researches, we could not draw on whole population study, but only on a large subsample. Second, one factor included only two indicators, which might lead to a lower reliability in the "Public identification as gay" factor. The data were collected in 2010 and it is likely that many Western societies have moved forward since then and become more accepting of gay/homosexual or bisexual men, perhaps limiting generalizability. Finally, the exclusion of the Spanish language sample (leaving only 8.3% of MSM living in Spain who completed the questionnaire in another language, and a small number of MSM in other countries who completed the questionnaire in Spanish) may have reduced the heterogeneity of the sample.

This is the first time multi-group models were adopted to investigate whether the factorial structure of IH was invariant across different populations of MSM, using the large EMIS sample. This study confirmed the validity of the SIHS in 38 European countries, with only minor differences across the three "homosexual discrimination" country clusters or across subgroups of age and education. In addition to conclusions in studies in the U.S. and in Uganda, the SIHS was shown to be appropriately stable in multi-populations in Europe. While this scale can be further investigated in other populations to identify the extent to which the SIHS is a standard measure for Internalized Homonegativity, it shows promise as a stable and psychometrically valid 7-item measure across homosexual discrimination levels, age and education.

Formatted: German

References

Bell, A. P., & Weinberg, M. S. (1978). *Homosexualities: A study of diversity among men and women*. NewYork: Simon & Schuster.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606. doi:10.1037/0033-2909.88.3.588

Berg, R. C., Ross, M., Weatherburn, P., & Schmidt, A. J. (2013). Structural and environmental factors are associated with internalised homonegativity in men who have sex with men: findings from the European MSM Internet Survey (EMIS) in 38 countries. *Social Science & Medicine*, *78*, 61-69. doi:10.1016/j.socscimed.2012.11.033

Berg, R. C., Weatherburn, P., Ross, M. W., Schmidt, A. J. (2015). The relationship of Internalized Homonegatvity to sexual health and wellbeing among men in 38 European countries who have sex with men. *Journal of Gay and Lesbian Mental Health*, *19*, 285-302. doi:10.1080/19359705.2015.1024375

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* NewYork: The Gulford Press.

Byrne, B. M. (2012). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming.* Oxfordshire, UK: Taylor & Francis Group, LLC.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measuement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037/0033-2909.105.3.456

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*, 287-321. doi:10.1207/s15328007sem1302_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255. doi:10.1207/S15328007SEM0902_5

Currie, M. R., Cunningham, E. G., & Findlay, B. M. (2004). The short internalized homonegativity scale: examination of the factorial structure of a new measure of internalized homophobia. *Educational and Psychological Measurement*, *64*, 1053-1067. doi:10.1177/0013164404264845

The EMIS Network (n.d.). *EMIS: the European MSM Internet Survey*. Retrieved from http://www.emis-project.eu/research-questions-and-covered-items.html

Finnegan, D., & Cook, D. (1984). Special issues affecting the treatment of gay male and lesbian alcoholics. *Alcoholism Treatment Quarterly*, *1*, 85-98. doi:10.1300/J020V01N03_07

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

Huebner, D. M., Davis, M. C., Nemeroff, C. J., & Aiken, L. S. (2002). The impact of internalized homophobia on HIV preventive interventions. *American Journal of Community Psychology*, *30*, 327-348. doi:10.1023/A:1015325303002

Lingiardi, V., Baiocco, R., & Nardelli, N. (2012). Measure of internalized sexual stigma for lesbians and gay men: a new scale. *Journal of Homosexuality*, *59*, 1191-1201. doi:10.1080/00918369.2012.712850

Formatted: French (Switzerland)

Margolies, L., Becker, M., & Jackson-Brewer, K. (1987). Internalized homophobia: identifying and treating the oppressor within. In Boston Lesbian Psychologies Collective (Ed.), *Lesbian Psychologies: Explorations and Challenges*, 229-241. Urbana, IL: University of Illinois.

Mayfield, W. (2001). The development of an internalized homonegativity inventory for gay men. *Journal of Homosexuality*, *41*, 53-76. doi:10.1300/J082v41n02_04

Meyer, I. H., & Dean, L. (1998). Internalized homophobia, intimacy, and sexual behavior among gay and bisexual men. In G. M. Herek (Ed.), *Stigma and sexual orientation: understanding prejudice against lesbians, gay men, and bisexuals* (pp. 160-186). Thousand Oaks, CA: Sage Publications.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological Research*, *3*, 111-121. doi:10.21500/20112084.857

Muthén, L., & Muthén, B. (2013). Mplus (Version 7.1). Los Angeles: Muthén & Muthén.

Nungesser, L. G. (1983). *Homosexual acts, actors and identities*. New York: Praeger Publishers.

O'Halleran Glaus, K. (1988). Alcoholism, chemical dependency and the lesbian client. *Women and Therapy*, *8*(1-2), 131-144.

Ross, M. W., Berg, R. C., Schmidt, A. J., Hospers, H. J., Breveglieri, M., Furegato, M., . . . European MSM Internet Survey (EMIS) Network. (2013). Internalised homonegativity predicts HIV-associated risk behavior in European men who have sex with men in a 38-country cross-sectional study: some public health implications of homophobia. *BMJ Open*, *3*, e001928. doi:10.1136/bmjopen-2012-001928

Ross, M. W., Kajubi, P., Mandel, J. S., McFarland, W., & Raymond, H. F. (2013). Internalized homonegativity/homophobia is associated with HIV-risk behaviours among Ugandan gay and

bisexual men. *International Journal of STD & AIDS*, *24*, 409-413. doi:10.1177/0956462412472793

Ross, M. W., Månsson, S-A., Daneback, K., Cooper, A. and Tikkanen, R. (2005). Biases in internet sexual health samples: comparison of an internet sexuality survey and a national sexual health survey in Sweden. *Social Science & Medicine*, *61*, 245-252. doi:10.1016/j.socscimed.2005.01.019

Ross, M. W., & Rosser, B. R. S. (1996). Measurement and correlates of internalized homophobia: a factor analytic study. *Journal of Clinical Psychology*, *52*, 15-21. doi:10.1002/(SICI)1097-4679(199601)52:1<15::AID-JCLP2>3.0.CO;2-V

Ross, M. W., Rosser, B. R. S., & Smolenski, D. (2010). The importance of measuring internalized homophobia/homonegativity. *Archives of Sexual Behavior*, *39*, 1207-1208. doi:10.1007/s10508-010-9634-z

Ross, M. W., Smolenski, D. J., Kajubi, P., Mandel, J. S., McFarland, W., & Raymond, F. H. (2010). Measurement of internalized homonegativity in gay and bisexual men in Uganda: cross-cultural properties of the Internalized Homonegativity scale. *Psychology, Health & Medicine*, *15*, 159-65. doi:10.1080/13548500903527746

Shidlo, A. (1994). Internalized homophobia: conceptual and empirical issues in measurement. In B. Greene & G. M. Herek (Eds.), *Lesbian and gay psychology: Theory, research, and clinical applications* (pp. 176-205). Thousand Oaks, CA: Sage Publications.

Shoptaw, S., Weiss, R. E., Munjas, B., Hucks-Ortiz, C., Young, S. D., Larkins, S., . . . Gorbach, P. M. (2009). Homonegativity, substance use, sexual risk behaviors, and HIV status in poor and ethnic

men who have sex with men in Los Angeles. *Journal of Urban Health*, *86*, 77-92. doi:10.1007/s11524-009-9372-5

Smolenski, D. (2009). *The application of latent variable models to the assessment of determinants of HIV risk behavior.* University of Texas: Houston, Texas.

Smolenski, D. J., Diamond, P. M., Ross, M. W., & Rosser, B. R. S. (2010). Revision, criterion validity, and multigroup assessment of the reactions to homosexuality scale. *Journal of Personality Assessessment*, *92*, 568-576. doi:10.1080/00223891.2010.513300

Stain, J. A., Lee, J. W., & Jones, P. S. (2006). Assessing cross-cultural differences through use of multiple-group invariance analyses. *Journal of Personality Assessment, 87*, 249-258. doi:10.1207/s15327752jpa8703_05

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 173-180. doi:10.1207/s15327906mbr2502_4

Weatherburn, P., Schmidt, A.J., Hickson, F., Reid, D., Berg, R. C., Hospers H. J., Marcus, U., & The EMIS Network (2013). The European Men who have sex with Men Internet Survey (EMIS): design and methods. *Sexuality Research and Social Policy*, 10(4), 243-257. doi:10.1007/s13178-013-0119-4

Weber, G. N. (2005). *Heterosexist events, internalized homophobia, and substance use among lesbian, gay, and bisexual individuals: Implications for counseling* (Unpublished doctoral dissertation). Penn State, University Park, PA.

Wikipedia. http://en.wikipedia.org/wiki/LGBQ+T_rights_by_country_or_territory, (accessed 30 Nov 2011).

Williamson, I. R. (2000). Internalized homophobia and health issues affecting lesbians and gay men. *Health Education Research*, 97-107. doi:10.1093/her/15.1.97

Table 1. Demographic characteristics of MSM with internalized homonegativity (IH) scores

Country	N	Median age	Percent with tertiary education	IH mean
"Liberal" country cluster				
Belgium	3341	34	66.6	1.357
Denmark	1422	34	50.9	1.250
Finland	1689	33	51.2	1.569
Ireland (Republic)	1838	31	71.3	1.598
Netherlands	3246	40	63.0	1.211
Norway	1720	31	66.0	1.320
Spain	1065	35	62.6	1.230
Sweden	2460	35	55.6	1.216
United Kingdom	15265	36	64.2	1.390
"Moderate" country cluster				
Austria	3339	31	31.1	1.310
Bosnia & Herzegovina	101	28	53.0	2.536
*Croatia**	409	29	64.6	2.104
Czech Republic	1891	27	42.1	1.550
France	9383	34	73.7	1.457
Estonia	487	30	48.0	1.703
Germany	44339	33	32.4	1.287
Greece	2282	30	65.6	1.976
Hungary	1546	28	68.9	1.678
Italy	12325	33	43.5	1.751
Luxembourg	226	36	50.7	1.356
Macedonia	87	28	80.5	2.218
Portugal	4011	31	59.7	1.777
Romania	1582	27	58.1	2.202
Slovenia	682	30	48.1	1.971
Switzerland	4165	37	46.7	1.365
Turkey	1297	28	85.6	2.383
"Conservative" country cluster				
Bulgaria	745	27	59.6	2.571
Belarus	293	27	68.6	2.267
Cyprus	212	30	64.5	2.239
Latvia	528	30	50.2	2.104
Lithuania	409	27	74.3	2.058
Malta	98	33	61.2	1.933
Moldova	86	24	59.3	2.510
Poland	1641	28	75.7	1.987
Serbia	813	28	57.3	2.239
Slovakia	433	26	50.2	1.702
Russia	3936	30	73.6	2.070
Ukraine	1296	28	76.1	2.195
Total	130718	33	50.4	1.498

Legend: *non-EU/EFTA countries* in italics; *Croatia was not part of the EU at the time of the survey

Table 2. Means, standard deviations, standardized coefficients, Spearman correlation of eight items of the short internalized homonegativity scale

Factor and Item	Mean	SD	β	S.E
Social comfort with gay man (SC)				
I1. Obviously effeminate homosexual men make me feel uncomfortable [R]	4.31	2.11	0.355	0.004
I2. I feel comfortable in gay bars	5.10	1.93	0.597	0.004
I3. Social situations with gay men make me feel uncomfortable [R]	2.77	2.03	0.456	0.003
Public identification as gay (PUBID)				
I4. I feel comfortable being seen in public with an obviously gay person	5.11	2.05	0.695	0.003
I5. I feel comfortable discussing homosexuality in a public situation	5.35	1.91	0.732	0.003
Personal comfort with a gay identity (PC)				
I6. I feel comfortable being a homosexual man	5.69	1.78	0.843	0.002
I7. Homosexuality is morally acceptable to me	6.49	1.32	0.576	0.003
I8. Even if I could change my sexual orientation, I wouldn't	5.55	2.01	0.843	0.003

	Correlation							
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**
I1. Obviously effeminate homosexual men make me feel uncomfortable [R]	-							
I2. I feel comfortable in gay bars	0.172	-						
I3. Social situations with gay men make me feel uncomfortable [R]	0.345	0.273	-					
I4. I feel comfortable being seen in public with an obviously gay person	0.284	0.285	0.339	-				
I5. I feel comfortable discussing homosexuality in a public situation	0.168	0.283	0.263	0.513	-			
I6. I feel comfortable being a homosexual man	0.172	0.397	0.300	0.462	0.501	-		
I7. Homosexuality is morally acceptable to me	0.113	0.249	0.204	0.330	0.349	0.442	-	
I8. Even if I could change my sexual orientation, I wouldn't	0.146	0.272	0.217	0.365	0.370	0.557	0.390	-

[R]: Reversed coded item; β: standardized coefficients of item with factor

Table 3. The multi-group validation in the confirmation split sample

8-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref Model	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
1. Model fitted into the training dataset		3607.731	15	1.260					0.960	0.926	0.062			
2. Model fitted into the testing dataset		3421.534	15	1.258					0.962	0.930	0.060			
A. Model freely estimated in both samples	Configural	7029.131	30	1.259					0.961	0.928	0.061			
B. All first order factor loadings constrained equal	Metric	6921.305	35	1.280	A	3.874	3	0.275	0.962	0.939	0.056	0.001	-0.005	Accept
C. All first order intercepts constrained equal	Scalar	7221.730	43	1.228	B	8.100	8	0.424	0.960	0.948	0.052	-0.002	-0.004	Accept
D. Residual variances constrained equal	Residual	6844.108	51	1.297	C	6.042	8	0.643	0.962	0.959	0.046	0.002	-0.006	Accept
E. All second factor loading constrained equal		6878.800	54	1.291	D	3.131	3	0.372	0.962	0.961	0.045	0.000	-0.001	Accept

7-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref Model	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
1. Model fitted into the training dataset		1409.627	12	1.290					0.982	0.969	0.043			
2. Model fitted into the testing dataset		1536.348	12	1.292					0.981	0.966	0.045			
A. Model freely estimated in both samples	Configural	2946.037	24	1.291					0.981	0.968	0.044			
B. Factor loadings constrained equal	Metric	2872.233	28	1.327	A				0.982	0.973	0.040	0.001	-0.004	Accept
C. Intercepts	Scalar	3027.807	35	1.261	B				0.981	0.977	0.037	-0.001	-0.003	Accept

constrained equal												
D. Residual variances constrained equal	Residual	2813.576	42	1.360	C		0.982	0.982	0.033	0.001	-0.004	Accept
E. Residual covariance constrained equal	Covariance residual	2820.068	43	1.357	D		0.982	0.983	0.032	0.000	-0.001	Accept

Note: $X^2_{S\text{-}B}$: Satorra-Bentler corrected chi-square; df: degrees of freedom; SCF: scaling correction factor; Ref Model: reference model for nested model testing; S-B ΔX^2: change in the Satorra-Bentler corrected chi-square; Δdf: change in the degree of freedom; p: p-value for the likelihood ratio test between nested models; CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean squared error of approximation

Table 4. Multi-group invariance testing by "Homosexual Discrimination" country clusters

8-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref Model	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
1. Model fitted into "liberal" countries		1522.454	15						0.967	0.939	0.057			
2. Model fitted "moderate" countries		4159.656	15						0.963	0.931	0.060			
3. Model fitted into "conservative countries"		1460.081	15						0.937	0.882	0.074			
A. Model freely estimated in both samples	Configural	7234.752	45	1.237					0.961	0.927	0.062			
B. First order factor loadings constrained equal	Metric	7390.294	55	1.266	A	290.819	10	<0.01	0.960	0.939	0.057	-0.001	-0.005	Accept
C. Intercepts constrained equal	Scalar	12282.035	71	1.207	B	5445.464	16	<0.01	0.934	0.922	0.064	-0.026	0.007	Reject
C1. Intercepts constrained equal except I2	Scalar	10911.001	69	1.213	B	3860.461	14	<0.01	0.941	0.928	0.062	-0.019	0.005	Reject
C2. Intercepts constrained equal except I2&I3	Scalar	9406.951	67	1.219	B	2103.428	12	<0.01	0.949	0.936	0.058	-0.011	0.001	Reject
C3. Intercepts constrained equal except I1&I2&I3	Scalar	8630.029	65	1.225	B	1216.288	10	<0.01	0.953	0.940	0.056	-0.007	-0.001	Accept
D. Residual variances constrained equal	Residual	11663.749	81	1.299	C3	2862.771	16	<0.01	0.937	0.935	0.059	-0.016	-0.003	Reject
D1. Residual variances constrained equal except I3	Residual	9840.657	79	1.309	C3	1359.410	14	<0.01	0.947	0.944	0.055	-0.006	-0.001	Accept
E. Second order factor loadings constrained equal		10045.361	85	1.300	D1	150.275	6	<0.01	0.946	0.947	0.053	-0.001	-0.002	Accept

7-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref Model	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
1. Model fitted into "liberal" countries		551.459	12						0.987	0.977	0.038			

2. Model fitted "moderate" countries		2182.963	12			0.978	0.961	0.049				
3. Model fitted into "conservative countries"		365.669	12			0.982	0.969	0.041				
A. Model freely estimated in both samples	Configural	3163.654	36	1.271		0.981	0.966	0.046				
B. Factor loadings constrained equal	Metric	3441.916	44	1.316	A	0.979	0.970	0.043	-0.002	-0.003	Accept
C. Intercepts constrained equal	Scalar	7491.711	58	1.240	B	0.954	0.950	0.056	-0.025	0.013	Reject
C1. Intercepts constrained equal except I2	Scalar	6046.675	56	1.249	B	0.963	0.958	0.051	-0.016	0.008	Reject
C2. Intercepts constrained equal except I2&I3	Scalar	4583.080	54	1.258	B	0.972	0.967	0.045	-0.007	0.002	Accept
D. Residual variances constrained equal	Residual	8010.779	68	1.363	C2	0.951	0.954	0.053	-0.021	0.008	Reject
D1. Residual variances constrained equal except I3	Residual	6320.861	66	1.382	C2	0.961	0.963	0.048	-0.011	0.003	Reject
D2. Residual variances constrained equal except I2&I3	Residual	5381.478	64	1.382	C2	0.967	0.968	0.045	-0.005	0.000	Accept
E. Residual covariance constrained equal	Covariance residual	5577.805	66	1.373	D2	0.966	0.967	0.045	-0.001	0.000	Accept

Note: $X^2_{S\text{-}B}$: Satorra-Bentler corrected chi-square; df: degrees of freedom; SCF: scaling correction factor; Ref Model: reference model for nested model testing; S-B ΔX^2: change in the Satorra-Bentler corrected chi-square; Δdf: change in the degree of freedom; p: p-value for the likelihood ratio test between nested models; CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean squared error of approximation; Accept: the nested model doesn't have lesser fit than the comparison model; Reject: the nested model has lesser fit than the comparison model.

Table 5. Multi-group invariance testing by age and education

8-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
Age (<33 vs. ≥33)														
1. Model fitted into MSM aged <33		3022.127	15						0.967	0.939	0.057			
2. Model fitted into MSM aged ≥33		4001.039	15						0.954	0.914	0.065			
A. Model freely estimated in both samples	Configural	7046.104	30	1.259					0.961	0.927	0.061			
B. First order factor loadings constrained equal	Metric	7226.036	35	1.282	A	276.573	5	<0.01	0.960	0.935	0.058	-0.001	-0.003	Accept
C. Intercepts constrained equal	Scalar	9489.832	43	1.230	B	2402.708	8	<0.01	0.947	0.931	0.059	-0.013	0.001	Reject
C1. Intercepts constrained equal except I4	Scalar	9036.818	42	1.235	B	1895.457	7	<0.01	0.950	0.933	0.059	-0.010	0.001	Reject
C2. Intercepts constrained equal except I4 &I8	Scalar	8570.255	41	1.241	B	1369.398	6	<0.01	0.952	0.935	0.058	-0.008	0.000	Accept
D. Residual variances constrained equal	Residual	8546.066	49	1.310	C2	336.410	8	<0.01	0.952	0.046	0.053	0.000	-0.005	Accept
E. Second order factor loadings constrained equal		8642.391	52	1.302	D	48.702	3	<0.01	0.952	0.948	0.052	0.000	-0.001	Accept
Education (between ISCED<5 & ISCED≥5)														
1. Model fitted into MSM with ISCED<5		3469.742	15						0.962	0.928	0.061			
2. Model fitted into MSM with ISCED≥5		3611.126	15						0.960	0.926	0.062			
A. Model freely estimated in both samples	Configural	7080.895	30	1.262					0.961	0.927	0.062			
B. First order factor loadings constrained equal	Metric	7264.072	35	1.282	A	268.504	5	<0.01	0.960	0.936	0.058	-0.001	-0.004	Accept
C. Intercepts constrained equal	Scalar	8466.449	43	1.229	B	1095.88	8	<0.01	0.953	0.939	0.056	-0.007	-0.002	Accept
D. Residual variances constrained equal	Residual	8242.652	51	1.298	C	175.982	8	<0.01	0.955	0.950	0.051	0.002	-0.005	Accept

E. Second order factor loading constrained equal	8372.748	54	1.292	D	99.688	3	<0.01	0.954	0.952	0.050	-0.001	-0.001	Accept	

7-item scale

Model	Validation type	X^2_{S-B}	df	SCF	Ref	S-B ΔX^2	Δdf	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Comment
Age (<33 vs. ≥33)														
1. Model fitted into MSM aged <33		1511.880	12						0.981	0.967	0.045			
2. Model fitted into MSM aged ≥33		1557.838	12						0.979	0.964	0.045			
A. Model freely estimated in both samples	Configural	3071.437	24	1.292					0.980	0.966	0.045			
B. Factor loadings constrained equal	Metric	3190.056	28	1.329	A				0.980	0.970	0.043	0.000	-0.002	Accept
C. Intercepts constrained equal	Scalar	5209.401	35	1.264	B				0.967	0.960	0.049	-0.013	0.006	Reject
C1. Intercepts constrained equal except I4	Scalar	4773.103	34	1.271	B				0.970	0.962	0.047	-0.010	0.004	Reject
C2. Intercepts constrained equal except I4 &I8	Scalar	4349.668	33	1.280	B				0.972	0.965	0.046	-0.008	0.003	Accept
D. Residual variances constrained equal	Residual	4514.097	40	1.379	C2				0.971	0.970	0.042	-0.001	-0.004	Accept
E. Residual covariance constrained equal	Covariance residual	4548.388	41	1.375	D				0.971	0.970	0.042	0.000	0.000	Accept
Education (ISCED<5 vs. ISCED≥5)														
1. Model fitted into MSM with ISCED<5		1538.630	12						0.981	0.967	0.046			
2. Model fitted into MSM with ISCED≥5		1402.111	12						0.982	0.968	0.043			
A. Model freely estimated in both samples	Configural	2941.675	24	1.293					0.981	0.967	0.044			
B. Factor loadings constrained equal	Metric	3015.038	28	1.329	A				0.981	0.971	0.041	0.000	-0.003	Accept
C. Intercepts constrained equal	Scalar	4064.546	35	1.263	B				0.974	0.969	0.043	-0.007	-0.002	Accept
D. Residual variances constrained equal	Residual	3949.886	42	1.362	C				0.975	0.975	0.039	0.001	-0.004	Accept
E. Residual covariance constrained	Covariance	3964.8	43	1.359	D				0.975	0.976	0.038	0.000	-0.001	Accept

equal		residual	84											

Note: X^2_{S-B}: Satorra-Bentler corrected chi-square; df: degrees of freedom; SCF: scaling correction factor; Ref Model: reference model for nested model testing; S-B ΔX^2: change in the Satorra-Bentler corrected chi-square; Δdf: change in the degree of freedom; p: p-value for the likelihood ratio test between nested models; CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean squared error of approximation; Accept: the nested model doesn't have lesser fit than the comparison model; Reject: the nested model has lesser fit than the comparison model; ISCED: the International Standardised Classification of Educational Degrees (0: Pre-primary education; 1: Primary education or first stage of basic education; 2: Lower secondary education or second stage of basic education; 3: Upper secondary education; 4: Post-secondary non-tertiary education; 5: First stage of tertiary education; 6: Second stage of tertiary education.

Table 6. Multi-group invariance testing by Homosexual *vs* Bisexual Identification

8-item scale

Model	**Validation type**	**X^2_{S-B}**	**df**	**SCF**	**Ref Model**	**S-B ΔX^2**	**Δdf**	**p**	**CFI**	**TLI**	**RMSEA**	**ΔCFI**	**ΔRMSEA**	**Comment**
1. Model fitted into the dataset including population self-defined as gays or homosexuals		4776.252	15	1.290					0.963	0.931	0.056			
2. Model fitted into the dataset including population self-defined as bisexuals		1610.466	15	1.139					0.912	0.836	0.086			
2.1. Model fitted into the dataset including population self-defined as bisexuals (correlation between Item and Item 3)		422.634	14	1.140					0.978	0.955	0.045			
A. Model freely estimated in both samples	Configural	5456.469	29	1.218					0.966	0.934	0.057			
B. All first order factor loadings constrained equal	Metric	6150.335	34	1.201	A	671.783	5	0.000	0.962	0.937	0.056	-0.004	-0.001	Accept
C. All first order intercepts constrained equal	Scalar	14481.666	42	1.171	B	9172.476	8	0.000	0.909	0.879	0.077	-0.053	0.021	Reject

C1. Intercept I1 constrained equal except I1	Scalar	6632.763	35	1.196	B	532.390	1	0.000	0.959	0.934	0.057	-0.003	0.001	Accept
D. Residual variances constrained equal	Residual	15589.186	43	1.162	C	10048.704	8	0.000	0.902	0.873	0.079	-0.057	0.020	Reject
D1. Residual variances constrained equal except I5,I6,I7;	Residual	8133.419	40	1.146	C	1743.861	5	0.000	0.949	0.929	0.059	-0.010	0.002	Accept

7-item scale

Model	**Validation type**	**X^2_{S-B}**	**df**	**SCF**	**Ref Model**	**S-B ΔX^2**	**Δdf**	**p**	**CFI**	**TLI**	**RMSEA**	**ΔCFI**	**ΔRMSEA**	**Comment**
1. Model fitted into the training dataset		5464.135	12	1.360					0.951	0.914	0.067			
2. Model fitted into the testing dataset		1540.732	12	1.141					0.906	0.836	0.094			
A. Model freely estimated in both samples	Configural	7348.924	24	1.250					0.948	0.908	0.073			
B. Factor loadings constrained equal	Metric	7734.622	28	1.233	A	310.021	4	0.000	0.945	0.917	0.069	-0.003	-0.004	Accept
C. Intercepts constrained equal	Scalar	16627.237	35	1.196	B	3325.530	7	0.000	0.881	0.858	0.091	-0.064	0.022	Reject
C1. Intercept I1 constrained equal except I1	Scalar	9131.545	29	1.226	B	1610.180	1	0.000	0.935	0.906	0.074	0.01	0.005	Accept
D. Residual variances constrained equal	Residual	16396.200	35	1.205	C	7759.082	6	0.000	0.883	0.859	0.090	-0.062	0.021	Reject
E. Residual	Covariance	9939.215	33	1.183	C	645.988	4	0.000	0.929	0.910	0.072	-0.006	-0.002	Accept

variances constrained equal except I6,I7.	residual										

Note: $X^2_{S\text{-}B}$: Satorra-Bentler corrected chi-square; df: degrees of freedom; SCF: scaling correction factor; Ref Model: reference model for nested model testing; S-B ΔX^2: change in the Satorra-Bentler corrected chi-square; Δdf: change in the degree of freedom; p: p-value for the likelihood ratio test between nested models; CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean squared error of approximation

Figure 1. The 8-item final second-order factor model developed by Smolenski, 2010.

Item 1: Obviously effeminate homosexual men make me feel uncomfortable (not used in the 7-items model); Item 2: I feel comfortable in gay bars; Item 3: Social situations with gay men make me feel uncomfortable; Item 4: I feel comfortable being seen in public with an obviously gay person; Item 5: I feel comfortable discussing homosexuality in a public situation; Item 6: I feel comfortable being a homosexual man; Item 7: homosexuality is morally acceptable to me; Item 8: even if I could change my sexual orientation, I wouldn't; SC: Social comfort with gay men; PUBID: Public identification as gay; PC: Personal comfort with a gay identity; IH: Internalized Homonegativity

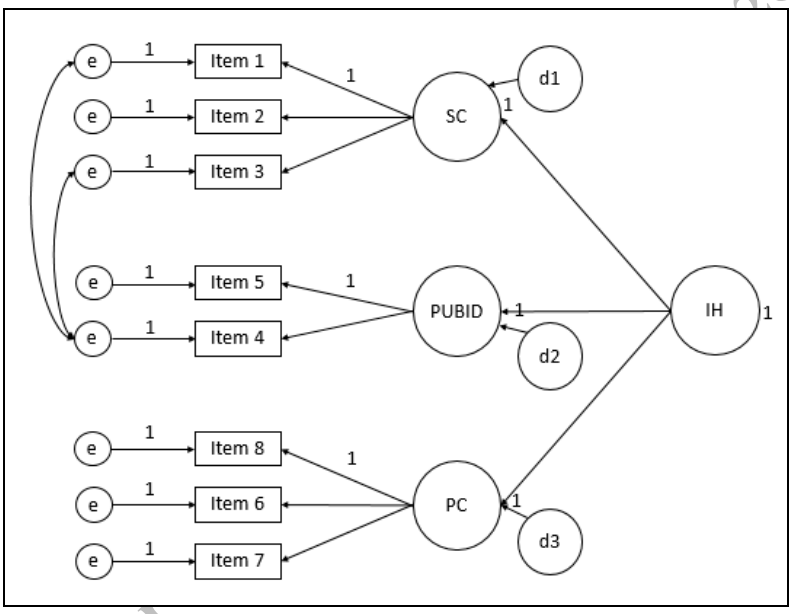

Figure 2: Parameter estimates from the final second-order factor model in the EMIS dataset.