# Measurement invariance of the General Health Questionnaire GHQ 12-item version (GHQ-12)

A large UK longitudinal study across students and non-students

**Citation for published version:**
Ushakova, A, McKenzie, K, Hughes, C, Stoye, J & Murray, AL 2023, 'Measurement invariance of the General Health Questionnaire GHQ 12-item version (GHQ-12): A large UK longitudinal study across students and non-students', *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000785

**Digital Object Identifier (DOI):**
10.1027/1015-5759/a000785

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
European Journal of Psychological Assessment

**Measurement Invariance of the General Health Questionnaire GHQ 12 item version (GHQ-12) Across Students and Non-Students based on a large UK Longitudinal Study.**

Anastasia Ushakova[ab], Karen McKenzie[c], Claire Hughes [d], Johanna Stoye[b], and Aja Louise Murray[b]

**Affiliations**: [a] Centre for Health Informatics, Computing and Statistics, Lancaster University, United Kingdom; [b]Department of Psychology, University of Edinburgh, United Kingdom; [c]Department of Psychology, Northumbria University, United Kingdom; [d]Department of Psychology, University of Cambridge, United Kingdom

**Author Note**

**Conflicts of interest/Competing interests:** The authors have no conflicts of interest relevant

to this article to disclose.

**Acknowledgements:** This work was funded by a **Student Mental Health Research Network (SMaRteN)** research grant. We are very grateful to all the families who took part in the Understanding Society Study, and the team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, and volunteers.

**Data and Code**: Data can be accessed via UKDS: https://ukdataservice.ac.uk/. The code that was used to conduct the analyses in the study is available at https://osf.io/jr6um/ .

**Measurement Invariance of the General Health Questionnaire GHQ 12 item version (GHQ-12) across Students and Non-Students based on a large UK Longitudinal Study.**


## Abstract

Understanding how levels, patterns, predictors, and outcomes of mental health issues differs in students relative to non-students can inform more effective and better tailored prevention and intervention for mental health in higher education contexts. However, comparisons of mental health in student and non-student groups depend on the critical but seldom-tested assumption of measurement invariance. In this study, we use data from the UK household longitudinal study (UKLS) to evaluate the measurement invariance of the scores from a commonly used mental health measure: the General Health Questionnaire 12-item version (GHQ-12) across students and non-students. Using a bifactor model to take account of wording factors we found measurement invariance up to the scalar level for students and non-student groups. This provides support for the use of instrument for comparing mental health issue levels and candidate risk factors and outcomes across students and non-students.



**Keywords:** measurement invariance; General Health Questionnaire; students; mental health

## 1. Introduction

Young people in higher education can be thought of as a particularly vulnerable group with respect to mental health. The transition to higher education, for example, comes with numerous academic, personal and social challenges and for a majority occurs at an age where there is already an elevated vulnerability for the onset of exacerbation of mental health issues (Andersen et al., 2021; Lewis et al., 2021). University and college students report high levels of mental health issues (Sheldon et al., 2021) and there is evidence that young people are reporting increasing levels of more serious mental health problems (HESA, 2018). To inform optimal prevention and intervention approaches it is important both to know whether students experience greater mental health difficulties than their peers and to identify risk factors and outcomes, including those that may be specific to student populations. These kinds of comparisons can help identify what is distinct about student mental health, enabling better tailoring of interventions to this population (Tabor et al., 2021).

Comparisons of student and non-student mental health; however, implicitly rely on the critical but seldom-tested assumption of measurement invariance across these groups. That is, one assumes that the same observed scores for students and non-students on measures of mental health assessments reflect the same underlying levels of a mental health issue (Svetina et al., 2020). Previous work, however, has suggested that students are at risk of perceiving both public and personal stigma surrounding mental health difficulties and that this negatively impacts help-seeking behaviour and reporting (Eisenberg et al., 2009; Martin, 2010; Shahwan et al., 2020). This, alongside other effects related to differential selection into and exposure to higher education environments could affect the way students interact with psychometric instruments when compared to non-students. For example, there could be under-reporting of some symptoms by students as compared to non-student peers with the same level of underlying mental health issue severity.

Different levels of invariance are required to support different types of comparisons across students and non-students (e.g., Murray, Speyer, et al., 2021). A common framework for understanding these is a confirmatory factor analysis latent variable framework. Within this framework, configural invariance describes the situation in which the same items can be used to measure an underlying construct across students and non-students. However, to compare variances and covariances across groups (e.g., comparing the associations between a candidate risk factor or outcome and mental health) metric invariance is required. This means that for ordinal items, the magnitude as well as the pattern of factor loadings is equivalent across students and non-students. To compare the levels (e.g., latent mean scores) of mental health constructs, scalar invariance is required (i.e., equality of both loadings and thresholds). Finally, to compare observed (as opposed to latent) scores across students and non-students in the case of ordinal items, residual invariance (i.e., equality of loadings, thresholds, and residual variances) is required.

Conversely, violations of invariance undermine student versus non-student comparisons. A lack of scalar invariance, for example, means that scores cannot be interpreted in the same way in students as in non-students and may result in invalid conclusions being drawn about differences in levels between students and non-students (Liu & West, 2018; Pokropek et al., 2019). For example, it is commonly noted that students experience higher levels of mental health issues than their non-student peers (e.g., Lewis et al., 2021); however, this is difficult to confirm without knowledge of whether the scores are comparable across these groups. Fortunately, it is often possible to obtain valid comparisons even when there are measurement invariance violations by modelling those violations (Pokropek et al., 2019). However, testing measurement invariance and identifying which parameters and in which items are non-invariant is a necessary step in this process. These non-invariant parameters can then be modelled as such within a latent variable model in order to avoid bias in the structural parameters that are used to compare groups (Pokropek et al., 2019).

Further, when invariance does not hold, the nature of the non-invariance can itself provide insights into differences between the two groups in how mental health symptoms are experienced and reported (e.g., Murray et al., 2021). For example, it could help to identify symptoms for which students or non-students are relatively less comfortable revealing, or flag items that may be less relevant or measured less reliably in one or other group.  It could also potentially reveal fundamentally different understandings of mental health among students and non-students (see e.g., Dodd et al., 2021). Altogether this could inform adaptations of measures for measuring mental health that are more suited for student populations as well as furthering our understanding of what is distinctive about student mental health.

One of the most popular measures used for the assessment of less severe psychological disorders, that can be used in non-clinical settings, is the General Health Questionnaire (GHQ) (Campbell et al., 2003; Doi & Minowa, 2003; Goldberg, 1972; Kalliath et al., 2004). Developed primarily in the UK, the measure is available in multiple forms (12, 20, 28, 30, and 60 items) and is used widely in psychological, epidemiological, and clinical contexts (Hankins, 2008). A recent scoping review demonstrated that variants of the GHQ are also commonly used in student mental health research (Dodd et al., 2021).

With the advantage of brevity, the 12-item version of the GHQ (the GHQ-12) is the most commonly used variants. Several researchers have conducted validation studies of the GHQ in various populations, such as clinical and non-clinical samples (Fernandes & Vasconcelos-Raposo, 2013), in adolescent populations in Australia (Tait et al., 2003), Japan (Doi & Minowa, 2003) and Ghana (Glozah & Pevalin, 2015). These studies generally conclude that their findings support the factorial validity and reliability of the GHQ-12.

There have also been psychometric studies of the GHQ-12 conducted in student samples. For example,  (Zulkefly & Baharudin, 2010) and  Lee & Kim (2020) fit a three-reported evidence for factorial validity of a three-factor model for the GHQ-12 and adequate reliability based on data from Malaysian and Korean students respectively. (Yaghubi et al., 2012) examined the factor structure, sensitivity, specificity, construct validity, and reliability of the GHQ-12 in a sample of medical students in Tehran. They reported that a two-factor structure was optimal and also concluded that their findings showed support for the other psychometric properties examined. However, as well as past research producing mixed findings on the optimal factor structure in student findings, we could identify no studies that examined the measurement invariance of the GHQ-12 across students and non-students.

Given the importance of testing measurement invariance for illuminating what is distinct about student mental health and the lack of studies in this area to date, the goal of the present study was to evaluate measurement invariance across students and non-students using a widely used measure of mental health: the GHQ-12.

## 2. Methods

### *2.1 Participants*

UK Household Longitudinal Study (UKHLS) is a longitudinal survey that covers approximately 100,000 individuals in over 40,000 households in the UK. The survey combines data from around 8,000 households from the British Household Panel Survey (BHPS), 1991-2009, and the Understanding Society Survey, 2009-Present. For the main analyses, a single wave (Wave 1 - 2009) data for Understanding Society Survey participants was used. Wave 1 provides the largest sample availability for the student and non-student groups, with data for ~3,000 and ~17,000 participants available for each group. Participants were invited annually to answer a series of questions including those that reveal whether they are currently in higher education. The variable *fenow* with the categories of '*Never been to college/university'* and '*At College/University'* was selected to represent contrasting groups for higher education attendance. Descriptive information regarding the student and non-student groups are provided in Tables 1 and 2.

More details on the dataset can be found on https://www.understandingsociety.ac.uk. The code that was used to conduct the analyses in the study is available at https://osf.io/jr6um/?view_only=4f35e9e8f91b472192954b28d22d08e2

### *2.2 Measures*

*General Health Questionnaire 12-item version.* The GHQ-12 includes 12 items and was originally designed to measure to measure a single unidimensional construct; however, items can also be labelled on the basis of measuring the sub-concepts of *Social Dysfunction* (6 items), *Anxiety* (4 items) and *Loss of Confidence* (2 items) (Lundin, 2016). Respondents rate their experience of each symptom in the past week using negatively worded questions, for example, '*Have you recently been thinking of yourself as a worthless person?'* Responses are recorded on a 4-point scale with higher scores representing poorer mental health.

### *2.3 Statistical Procedure*

To provide evidence of the measurement equivalence across the student and non-student groups, a confirmatory factor analysis model approach was used. Though the measure was originally proposed to be unidimensional, alternative structures have been proposed, some of which include wording factors to account for artefactual multidimensionality due to the presence of both positively and negatively worded items (see e.g., for an overview Gnambs & Staufenbiel, 2018). In brief, past literature has also provided supporting evidence for a unidimensional model (Banks & Jackson, 1982; Winefield et al., 1989), and a 2-factor model (Politi et al., 1994). In their large-N meta-analytic study comparing different structures (Gnambs & Staufenbiel, 2018) recommended a bifactor model with positively and negatively worded group factors to account for wording effects. We, therefore, adopted this structure as the basis for our analyses in the present study. However, we also fit and compared a unidimensional model and oblique 2- and 3-factor models for comparison to check whether their proposed bifactor model also captured the item covariances best in the present sample. We did this for both the student and non-student sub-samples.

The one-factor model loaded all items on a single dimension. In the two-factor model items 1,3,4,7,8, and 12 formed one dimension while the remaining items formed the other (Gnambs & Staufenbiel, 2018). In the three-factor model: items 1,3,4,7,8 and 12 loaded on the first factor; items 2,5,6, and 9 loaded on the second and items 10 and 11 loaded on the third. Finally, in the bifactor model with wording factors, all items loaded on a general factor, items 1,3,4,7,8 and 13 loaded on specific factor 1, and items 2,5,6,9,10,11 loaded on factor 2.

If the same factor structure was supported in both groups, we judged configural invariance to hold and we proceeded to test metric and scalar invariance. Given the ordered-categorical nature of the scale (<5 response options), ordinal data measurement invariance (MI) procedures were used in line with the recommendation by Svetina et al. (2020). A series of incremental models were implemented to test for invariance, starting with the baseline model with no constraints on threshold and loadings across the groups, then adding threshold constraints and finally loading constraints. All analyses were performed using R. The implementation was directly guided by recommendations of Svetina et al. (2020) for multi-group invariance analyses in the ordinal setting and conducted in *lavaan* (Rosseel, 2012) in R statistical software.

The first step of the evaluation of measurement invariance is a setup of the baseline model, where number and patterns of the key parameters are assumed to be equal across groups with threshold and loadings values being allowed to vary, except for minimal cross-group constraints needed for model identification. The baseline model was specified using ordinal representation for the items with the delta parameterization for model specification (Wu & Estabrook, 2016).

Various approaches  can be used to then evaluate the invariance and the optimal approach may depend on the number of factors, the number of groups, and the size of the groups that are being compared (Svetina et al., 2020). In the present study we adopted the criteria of Chen (2007), which is based on a comprehensive simulation study. For the group sample sizes in the present study, these criteria are that metric invariance holds if the addition of metric constraints (here threshold constraints are added first following (Svetina et al., 2020) lead to a decrease in CFI of no more than  .010, supplemented by an increase of no more than.015 in RMSEA and .030 in SRMR. Scalar invariance holds if the addition of scalar constraints (here adding loading constraints to the threshold constraint) leads to a decrease of no more than .010 in CFI, increase of no more than .015 in RMSEA and .010 in SRMR. Chi-square difference tests are also reported for information; however, it has been well-documented that these can be overly sensitive to mis-specification in an invariance testing context (Yuan & Chan, 2016). We, therefore, do not use these as a basis for judging if invariance holds.

## 2. Results

### Descriptive statistics

Descriptive statistics that summarise the distribution of responses for each GHQ-12 items for each group are provided in Table 2.

### Single group CFAs

The model fits for the single group CFAs are provided in Table 2. These suggested that the bifactor model was the best fitting in both the student and non-student groups. Given that this is consistent with the recommended model from a recent meta-analytic study, we adopted this model for our measurement invariance analyses (Gnambs & Staufenbiel, 2018). We selected this model despite the fact that superior bifactor model fit may sometimes reflect the presence of methodological artefacts (Murray & Johnson, 2013) because previous research has suggested the presence of wording variance in the GHQ-12 that can be accounted for with the bifactor model (Gnambs & Staufenbiel, 2018). That is, in this case a source of methodological artefact has been identified and is appropriately modelled with a bifactor model.

### Measurement invariance analyses

The model fits for each level of invariance testing are provided in Table 3. The baseline (configural) model fit well. The addition of threshold constraints led to a deterioration in fit which was statistically significant according to a chi-square difference test [delta-chi-square (12) = 73.419, $p<.01$]. However, the deterioration in fit was within the bounds acceptable by Chen's (2007) criteria and it was concluded that invariance held at this level. The addition of loading constraints to this model then led to an improvement in model fit overall, though the chi-square difference test was also significant here [delta-chi-square (21) = 55.881, $p<.01$]. As such, scalar invariance was judged to hold. This was taken as our final model. The loading parameter estimates for this model are provided in Table 4. These are suggestive of a strong general factor (loadings |.46 |- |.90|; omega hierarchical = .86; explained common variance =.79) but relatively weak specific factors (loadings |.08| - |.57|, omega hierarchical for S1 = .33, S2 = .01).

## Discussion

To provide a robust foundation for valid comparisons of student and non-student mental health levels, risk factors, and outcomes we conducted a measurement invariance analysis of the GHQ-12 across student and non-student groups in a large population-representative sample. Results suggested that measurement invariance held for a bi-factor factor model of the GHQ-12 up to the scalar level. This supports the validity of using the GHQ-12 to compare mental health and predictive risk factors (and outcomes) across students and non-students within latent variable measurement models. It thus provides a critical foundation for illuminating differences in levels, risk factors for, and outcomes of, mental health issues in students compared to non-students (Tabor et al., 2021).

In line with previous research we also found that a bi-factor model fit well to the GHQ-12 data and confirmed that this was the case for both students and non-students (Gnambs & Staufenbiel, 2018). This is in contradiction to some previous studies in student samples that advocated for a three-factor structure (e.g., Lee & Kim, 2020; Zulkefly & Baharudin, 2010). However, these studies did not address the possibility of wording effects nor provide a direct comparison of the three-factor model with a bi-factor model. Further, also consistent with previous research (Hystad et al., 2020), we found that the general factor was strong and the specific factors were weak and unreliable in terms of what they added over and above the measurement of the general factor. This suggests that comparisons of students and non-students could use a bifactor model to ensure that the wording effects are accounted for but focus on differences in the levels of, risk factors for, and outcomes of the general latent variable.

Knowledge of these differences is important for understanding how to tailor intervention and prevention to students. For example, robust knowledge of which mental health issues are most elevated in students *versus* non-students and whether risk factors established in the general population have the same importance in students can help optimise the provision of support and identifying targets for interventions aimed at students that complement mental health provision available to the general population. While several previous studies have compared student and non-student mental health (Blanco et al., 2008; Tabor et al., 2021), none to the best of our knowledge has yet done so ensuring that observed differences did not merely reflect differences in the way that items are understood or responded to in students versus non-students. Given our finding of scalar invariance in the GHQ-12, our results suggest that this instrument represents a good choice of measure for future studies that seek to illumine student versus non-student differences.

It is important to note that these findings pertain to the use of a latent variable model as scalar invariance provided unbiased comparisons only for latent means. Stricter invariance (up to the residual level) is required for comparisons based on observed scores. However, given that latent variable measurement models can provide more reliable measurement of the underlying constructs and also can be used to model the wording effects that have been identified in the GHQ-12, it is advisable that a latent variable measurement model be used for student and non-student comparisons in any case.

**Limitations and Future Directions**

The scope of our study relates to UK students and further research could test the multi-group invariance in various regions across the world. Further, the GHQ is available in various forms. Our results would be valid for GHQ-12 format but would not necessarily hold in other versions such as GHQ-28, GHQ-30, or GHQ-60. Further research to investigate other forms of the instrument would be needed. Methodologically, to complement the analysis, other notable methods might be considered for multi-group invariance analysis to strengthen the evidence presented by this research. These include the alignment method of and among others, Bayesian extensions: Bayesian structural equation modelling (SEM) and partial multigroup Bayesian SEM. Both were recently surveyed and compared to multigroup factor analysis approaches in (Pokropek et al., 2019). Furthermore, to provide a more comprehensive picture of the validity of mental health measures more generally for this and similar datasets, a natural extension would be to assess whether other commonly used wellbeing measures such as the Warwick Edinburgh Mental Wellbeing Scale (WEWMBS) exhibit invariance across students and non-students. Finally, it would also be helpful to examine longitudinal by group invariance of the GHQ-12 in future studies. This would facilitate valid comparisons not only of levels, risk factors, and outcome across students and non-students, but valid comparisons of their mental health trajectories over (Murray et al., 2017). This is important for, as an example, understanding the effects of transitions into and out of higher education.

## Conclusions

Configural, metric, and scalar invariance held across students and non-students' groups in the large UK-representative longitudinal survey, supporting the use of scores from this measure to investigate differences in the levels, risk factors, and outcomes of mental health across these groups. Further research may consider replicating the research in other countries and extend the analyses to the assessment of group-by-longitudinal invariance across both groups to support comparisons of mental health trajectories (and their predictors and outcomes) in students versus non-students.

## References


Andersen, R., Holm, A., & Côté, J. E. (2021). The student mental health crisis: Assessing psychiatric and developmental explanatory models. *Journal of Adolescence*, *86*, 101–114.

Banks, M. H., & Jackson, P. R. (1982). Unemployment and risk of minor psychiatric disorder in young people: Cross-sectional and longitudinal evidence. *Psychological Medicine*, *12*(4), 789–798.

Blanco, C., Okuda, M., Wright, C., Hasin, D. S., Grant, B. F., Liu, S.-M., & Olfson, M. (2008). Mental health of college students and their non–college-attending peers: Results from the national epidemiologic study on alcohol and related conditions. *Archives of General Psychiatry*, *65*(12), 1429–1437.

Campbell, A., Walker, J., & Farrell, G. (2003). Confirmatory factor analysis of the GHQ-12: Can I see that again? *Australian & New Zealand Journal of Psychiatry*, *37*(4), 475–483.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Dodd, A. L., Priestley, M., Tyrrell, K., Cygan, S., Newell, C., & Byrom, N. C. (2021). University student well-being in the United Kingdom: A scoping review of its conceptualisation and measurement. *Journal of Mental Health*, *30*(3), 375–387.

Doi, Y., & Minowa, M. (2003). Factor structure of the 12-item General Health Questionnaire in the Japanese general adult population. *Psychiatry and Clinical Neurosciences*, *57*(4), 379–383.

Eisenberg, D., Downs, M. F., Golberstein, E., & Zivin, K. (2009). Stigma and help seeking for mental health among college students. *Medical Care Research and Review*, *66*(5), 522–541.

Fernandes, H. M., & Vasconcelos-Raposo, J. (2013). Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment*, *20*(2), 219–229.

Glozah, F. N., & Pevalin, D. J. (2015). Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *Journal of Child & Adolescent Mental Health*, *27*(1), 53–57.

Gnambs, T., & Staufenbiel, T. (2018). The structure of the General Health Questionnaire (GHQ-12): Two meta-analytic factor analyses. *Health Psychology Review*, *12*(2), 179–194.

Goldberg, P. (1972). The detection of psychiatric illness by questionnaire. *Maudsley Monograph*.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): The result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, *4*, 1–8.

Hesa, H. (2018). Higher education student statistics: UK, 2016/17. *HESA, Promenade*.

Kalliath, T. J., O'Driscoll, M. P., & Brough, P. (2004). A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *20*(1), 11–20.

Lee, B., & Kim, Y. (2020). Factor structure of the 12-item General Health Questionnaire (GHQ-12) among Korean university students. *Psychiatry Clin. Psychopharmacol*, *30*(1).

Lewis, G., McCloud, T., & Callender, C. (2021). *Higher education and mental health: Analyses of the LSYPE cohorts: Research report: May 2021*.

Liu, Y., & West, S. G. (2018). Longitudinal measurement non-invariance with ordered-categorical indicators: How are the parameters in second-order latent linear growth models affected? *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 762–777.

Martin, J. M. (2010). Stigma and student mental health in higher education. *Higher Education Research & Development*, *29*(3), 259–274.

Murray, A. L., Hemady, C. L., Dunne, M., Foley, S., Osafo, J., Sikander, S., Madrid, B., Baban, A., Taut, D., & Ward, C. (2021). *Measuring antenatal depressive symptoms across the world: A validation and cross-country invariance analysis of the Patient Health Questionnaire–9 (PHQ-9) in eight diverse low resource settings*.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422.

Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2017). Evaluating longitudinal invariance in dimensions of mental health across adolescence: An analysis of the Social Behavior Questionnaire. *Assessment*, 1073191117721741.

Murray, A. L., Speyer, L. G., Hall, H. A., Valdebenito, S., & Hughes, C. (2021). A Longitudinal and Gender Invariance Analysis of the Strengths and Difficulties Questionnaire Across Ages 3, 5, 7, 11, 14, and 17 in a Large UK-Representative Sample. *Assessment*, 10731911211009312.

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A monte carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 724–744.

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, *90*(6), 432–437.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1–36.

Shahwan, S., Lau, J. H., Goh, C. M. J., Ong, W. J., Tan, G. T. H., Kwok, K. W., Samari, E., Lee, Y. Y., Teh, W. L., & Seet, V. (2020). The potential impact of an anti-stigma intervention on mental health help-seeking attitudes among university students. *BMC Psychiatry*, *20*(1), 1–14.

Sheldon, E., Simmonds-Buckley, M., Bone, C., Mascarenhas, T., Chan, N., Wincott, M., Gleeson, H., Sow, K., Hind, D., & Barkham, M. (2021). Prevalence and risk factors for mental health problems in university undergraduate students: A systematic review with meta-analysis. *Journal of Affective Disorders*, *287*, 282–292.

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semTools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 111–130.

Tabor, E., Patalay, P., & Bann, D. (2021). Mental health in higher education students and non-students: Evidence from a nationally representative panel study. *Social Psychiatry and Psychiatric Epidemiology*, *56*(5), 879–882.

Tait, R. J., French, D. J., & Hulse, G. K. (2003). Validity and Psychometric Properties of the General Health Questionnaire-12 in Young Australian Adolescents. *Australian & New Zealand Journal of Psychiatry*, *37*(3), 374–381. https://doi.org/10.1046/j.1440-1614.2003.01133.x

Winefield, H. R., Goldney, R. D., Winefield, A. H., & Tiggemann, M. (1989). The General Health Questionnaire: Reliability and validity for Australian youth. *Australian and New Zealand Journal of Psychiatry*, *23*(1), 53–58.

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045.

Yaghubi, H., Karimi, M., & Omidi, A. (2012). Validity and Factor Structure f the General Health Questionnaire (GHQ-12) In University Students. *International Journal Of Behavioral Sciences-□□□□ □□□□□□□ □□□□*, *6*(2).

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods*, *21*(3), 405.

Zulkefly, N. S., & Baharudin, R. (2010). Using the 12-item General Health Questionnaire (GHQ-12) to assess the psychological health of Malaysian college students. *Global Journal of Health Science*, *2*(1), 73.

**Table 1: Descriptive statistics**

| **Higher education status** | Female | Male | Total |
|---|---|---|---|
| *Never been to college/university* | 9811 (57%) | 7424 (43%) | 17235 |
| *At college/university* | 1676 (54%) | 1404 (46%) | 3080 |

**Table 2: Item category distributions for students and non-students**

| Item | Response category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Students** | | | | **Non-students** | | | |
| | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** |
| **1** | 289(9%) | 2347(76%) | 398(13%) | 46(1%) | 690 (4%) | 13680(79%) | 2453 (14%) | 412 (2%) |
| **2** | 1170(38%) | 1317(43%) | 481(16%) | 112(4%) | 5560(32%) | 8538 (49%) | 2374 (14%) | 763 (4%) |
| **3** | 429(14%) | 2308(75%) | 294(10%) | 49(2%) | 1401(8%) | 13541(79%) | 1780(10%) | 513 (3%) |
| **4** | 477(15%) | 2352(76%) | 232(8%) | 19(1%) | 1143 (7%) | 14419(84%) | 1444 (8%) | 229(1%) |
| **5** | 868(29%) | 1276(43%) | 686(23%) | 150(5%) | 4603 (27%) | 8795 (51%) | 3107(18%) | 730(4%) |
| **6** | 1175(38%) | 1448(47%) | 382(12%) | 75(2%) | 6184 (36%) | 8665(50%) | 1861(11%) | 525(3%) |
| **7** | 420(14%) | 2183(71%) | 408(13%) | 69(2%) | 920 (5%) | 12911(75%) | 2712 (16%) | 692 (4%) |
| **8** | 440(14%) | 2338(76%) | 253(8%) | 49(2%) | 1117 (6%) | 14032(81%) | 1709(10%) | 377(2%) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **9** | 1308(42%) | 1139(37%) | 512(17%) | 121(4%) | 7080(41%) | 6524(38%) | 2829(16%) | 802(5%) |
| **10** | 1588(52%) | 990(32%) | 417(14%) | 85(3%) | 8185(47%) | 6158(36%) | 2269(13%) | 623(4%) |
| **11** | 2210(72%) | 612(20%) | 219(7%) | 39(1%) | 11632(67%) | 3974(23%) | 1188(7%) | 441(3%) |
| **12** | 532(17%) | 2187(71%) | 316(10%) | 45(1%) | 1592(9%) | 13468(78%) | 1772(10%) | 403(2%) |

**Table 3: Comparison of proposed factor models for the GHQ-12 in the student and non-student samples**

| | Chi-square | df | *P* | CFI | TLI | RMSEA |
|---|---|---|---|---|---|---|
| | | | **Students** | | | |
| Single factor | 3169.836 | 54 | <.01 | 0.883 | 0.857 | 0.137 |
| Two-factor | 1182.147 | 53 | <.01 | 0.958 | 0.947 | 0.083 |
| Three-factor | 979.905 | 51 | <.01 | 0.965 | 0.955 | 0.077 |
| Bifactor | 546.381 | 42 | <.01 | 0.981 | 0.970 | 0.062 |
| | | | **Non-students** | | | |
| Single factor | 19098.279 | 54 | <.01 | 0.929 | 0.913 | 0.143 |
| Two-factor | 6445.613 | 53 | <.01 | 0.976 | 0.970 | 0.084 |
| Three-factor | 5196.035 | 51 | <.01 | 0.981 | 0.975 | 0.077 |
| Bifactor | 2668.823 | 42 | <.01 | 0.990 | 0.985 | 0.060 |

**Table 3**

**Measurement Invariance Across Students and Non-Students**

| Model | $\chi2$ | Df | *p* | RMSEA | CFI | TLI | SRMR |
|---|---|---|---|---|---|---|---|
| Baseline Model | 2999.281 | 84 | <.01 | 0.058 | 0.991 | 0.985 | 0.026 |
| Threshold Invariance | 3077.804 | 96 | <.01 | 0.055 | 0.990 | 0.990 | 0.026 |
| Threshold and Loadings Invariance | 2456.846 | 117 | <.01 | 0.044 | 0.998 | 0.991 | 0.026 |

*Note*. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI= Tucker–Lewis index; SRMR= Standardised Root Mean Residual. *p < .05. **p < .01.

**Table 4: Factor loading parameters for final model**

| | Estimate | SE | *p* | Estimate | SE | *p* | Estimate | SE | *p* |
|---|---|---|---|---|---|---|---|---|---|
| | **General factor** | | | **Group factor 1** | | | **Group factor 2** | | |
| **GHQ1** | 0.600 | 0.005 | <.001 | 0.426 | 0.006 | <.001 | - | - | - |
| **GHQ2** | 0.731 | 0.005 | <.001 | - | - | - | 0.265 | 0.01 | <.001 |
| **GHQ3** | 0.458 | 0.006 | <.001 | 0.469 | 0.007 | <.001 | - | - | - |
| **GHQ4** | 0.509 | 0.006 | <.001 | 0.569 | 0.007 | <.001 | - | - | - |
| **GHQ5** | 0.794 | 0.005 | <.001 | - | - | - | 0.394 | 0.012 | <.001 |
| **GHQ6** | 0.785 | 0.004 | <.001 | - | - | - | 0.13 | 0.009 | <.001 |
| **GHQ7** | 0.609 | 0.005 | <.001 | 0.404 | 0.006 | <.001 | - | - | - |
| **GHQ8** | 0.589 | 0.005 | <.001 | 0.504 | 0.006 | <.001 | - | - | - |
| **GHQ9** | 0.876 | 0.003 | <.001 | - | - | - | 0.08 | 0.009 | <.001 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GHQ10** | 0.897 | 0.003 | <.001 | - | - | - | -0.173 | 0.011 | <.001 |
| **GHQ11** | 0.863 | 0.004 | <.001 | - | - | - | -0.237 | 0.012 | <.001 |
| **GHQ12** | 0.674 | 0.004 | <.001 | 0.329 | 0.006 | <.001 | - | - | - |