A Mokken Scale Analysis of the Kessler-6 Screening Measure among Chinese Elderly Population: Findings from a national survey

Lisong Zhang
Nanjing University of Posts and Telecommunications
Zhongquan Li (zqli@nju.edu.cn)
Nanjing University https://orcid.org/0000-0002-8480-4166

Research article

Keywords: Psychological distress, Mokken scale analysis, dimensionality, Differential item functioning, gender difference

Posted Date: May 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23181/v1

Version of Record: A version of this preprint was published on September 22nd, 2020. See the published version at https://doi.org/10.1186/s12877-020-01771-w.

Abstract

Background

The ageing population increases rapidly across the world. Timely and effective screening of their mental health problems is important to individuals, families and the whole society. The Kessler-6 screening measure (K6) is a very popular instrument for non-specific psychological distress. However, few studies have focused on the psychometric properties of this instrument in the elderly population.

Methods

The present study employed Mokken scale analysis to evaluate its dimensionality and structure, and employed differential item functioning (DIF) to examine whether the same structure existed across sex in a national representative sample of old Chinese people. Data were drawn from a public data set, the 2010 China Family Panel Studies (CFPS2010), and responses from a total of 6450 participants aged 60 years old and above (3136 males and 3314 females) were included in the final analysis.

Results

Mokken scale analysis supported the unidimensional structure of the K6. Differential item functioning (DIF) analysis revealed that two of the six items ("Hopeless" and "Everything was an effort") were marked for DIF based on the Chi-square. However, their impacts were negligible in term of McFadden's pseudo R^2.

Conclusions

The K6 demonstrates adequate psychometric properties in the old Chinese population. The sum of all six items can be used as an indicator of non-specific psychological distress. Differences in the indicator across gender should be considered as a real difference in psychological distress between the female and the male.

Background

The ageing population increases rapidly across the world. Mental-health problems such as depression and anxiety are prevalent in this population. They have both short-term and long-term consequences to individuals, families and the whole society [1]. According to Report on National Mental Health Development in China (2017–2018), in the past several years, prevalence estimate of depression disorder is ranged from 15–39.86%, and prevalence rate of anxiety disorder is ranged from 11.51–22.02% among Chinses elderly population[2]. Another survey with a large nationally representative sample (the China Health and Retirement Longitudinal Study (CHARLS)) also indicated that about 33.09% of Chinese older adults suffered depression disorders [3]. In consideration of the largest population and fastest aging in

China[4], timely and effective screening of psychological distress is vital to help these in risk for early intervention.

The 6-item version of the Kessler Psychological Distress Scale (K6), a very brief instrument, has been developed to screening for non-specific psychological distress[5]. It was initially designed for fast and accurate detection of severe mental illness among the general population. Later, it is also used in some clinical situations [6]. It demonstrates strong psychometric properties in many populations, such as emerging adolescents[7], adults[8], and the elderly[9]. It even outperforms the K10, a long-form with ten items, in screening for DSM-IV mood or anxiety disorder[10]. Due to its excellent performance and high efficiency, it is widely employed in several major global and national surveys, such as the WHO World Mental Health (WMH) Survey, the US National Health Interview Survey [6], the Australian National Survey of Mental Health and Well-Being[10], the Canadian National Population Health Survey[11], the South African Stress and Health study[12]. It is also included in the China Family Panel Studies (Institute of Social Science Survey, Peking University, 2015), a longitudinal survey of Chinese communities, families, and individuals.

However, still some debates exist about the dimensionality of the K6, which is critical in interpreting scores on the scale. The K6 was initially developed as a measure for a unidimensional construct[5]. The one-factor solution (with all items loading on a single factor) is also supported in most of the current studies [6, 8, 11, 13–20]. Nevertheless, this model had a poor fit with the data from a large sample of adolescents in Australia, and a modified single-factor model was proposed instead[21], allowing residuals correlations among several items. Moreover, two-factor models were also reported in several studies [6, 7, 22, 23]. Kessler et al. found a two-factor solution emerged in the Indian sample, with an item ("Everything was an effort") loading on the second factor[6]. Lee et al. examined the dimensionality of the K6 among 3014 Hong Kong residents[22]. They found a two-factor model best fit the data, with three items("Nervous", "Restless or fidgety", and "Everything was an effort") loading on the anxiety factor, another three items("Hopeless", "Depressed", and "Worthless") loading on the depression factor. Bessaha compared several models of the K6 among a large sample of emerging adults in the US and revealed that a two-factor model and a second-order two-factor model fit the data better than a one-factor model[7]. In their two-factor model, two items ("Nervous" and "Restless or fidgety") loaded on the anxiety factor, while the other four items loaded on the depression factor. Moreover, the anxiety factor and the depression factor loaded on psychological distress in the second-order two-factor model. Easton et al. reported a better fit of Bassaha's two-factor model than the unidimensional model to the responses from Palestinian social workers[23].

Traditionally, exploratory factor analysis(EFA) and confirmatory factor analysis(CFA) are used in examing the factor structure of the K6[11, 21]. Mokken scale analysis (MSA) has demonstrated its unique value in addressing the problem of dimensionality [24–26]. It belongs to the family of nonparametric item response theories. It assumes that all items in a scale are hierarchically ordered along the continuum of a latent construct. It is more flexible than IRT models like the Rasch model and logistic models regarding statistical assumptions and sample size. It is also superior to traditional factorial analysis in evaluating

dimensionality and model simultaneously, and suffering less distortion from item-score distribution [25]. Monotone homogeneity model (MHM), the most general Mokken model, has assumptions of unidimensionality, local independence, and latent monotonicity. Unidimensionality implies that all items measure the same latent construct. Mokken scale analysis provides an automated item selection procedure (AISP) to help assess the latent structure of a scale [27, 28]. The total score of all items reveals different levels of the latent construct [25]. The first aim of our study is to employ Mokken scale analysis to evaluate the dimensionality of the K6.

The K6 is often used in the comparison of psychological distress across ages, sex, education, job categories, and nations [6, 11, 20, 29]. Most studies implicitly assume that the K6 measures psychological distress in the same way in different groups. However, the assumption is not always correct and should be justified before comparison [30]. Regarding the findings on the K6, women are higher than men in mean level and prevalence of psychological distress in both adolescent and adult populations [21]. The differences may be the results of higher vulnerability and more exposure to stress events for women, or the consequences of the way they understand some items [11]. Few studies have examined measurement invariance for the K6 across gender, with the exceptions of Drapeau et al. and Mewton et al. [11, 21]. Drapeau et al. used multi-group confirmatory factor analyses to test gender invariance in different age groups with data from the Canadian National Population Health Survey. They found though some items might vary over life-course in the gender invariance patterns, the K6 hold measurement invariance across gender in general. Mewton et al. also examined gender invariance in a sample of Australian adolescents under the framework of confirmatory factor analysis. They indicated that the data didn't support the strong invariance model, and further examination of partial invariance models suggested that all items lack invariance in the thresholds. These studies are conducted among people of different ages in western cultures.We are not sure whether the findings could be replicated in eastern cultures, such as China.

Multi-group confirmatory factor analysis is commonly used in examining measurement invariance, but it might be not accurate in figuring out the source of non-invariance. A flexible and robust iterative hybrid logit regression/ item response theory (LR/IRT) framework is recommended to deal with such a problem[31]. The logit regression approach makes comparisons among several models representing the prediction of latent trait and group membership on item performance. In the same time, item response theory (IRT) models provide the estimation of latent trait scores. Simulation studies have proved the advantage of this framework in detecting DIF in comparison to other methods. Therefore, the second aim of our study is to employ differential item functioning (DIF) analysis to evaluate measurement invariance of the K6 across gender.

In all, the present study would investigate the dimensionality of the K6 and its measurement invariance across gender with data from China Family Panel Studies in the year 2010. The results would contribute to the understanding of the factor structure of the K6 in eastern cultures, and shed some light on the gender difference in psychological distress.

Methods

Data and Sample

The study is based on secondary data analysis. The data was drawn from a publicly available dataset, the 2010 China Family Panel Studies (CFPS2010). The CFPS was launched in 2010, by the Institute of Social Science Survey of Peking University, with financial support from the Chinese government. It is an annual longitudinal survey of Chinese national representative communities, families and individuals. It collects information related to a variety of topics, such as economic activities, education outcomes, family dynamics and relationships, migration, and health (Institute of Social Science Survey, Peking University, 2015). In the CFPS2010, a subsample of 6598 records from older adults (aged 60 and above) was selected. Due to missing responses to any K6 item, 148 records were discarded. Therefore, responses from a total of 6450 participants (3136 males and 3314 females) were included in the final analysis. Their age ranged from 60 to 110 years old, with a mean of 68.51 years (SD = 6.94).

Measures

The K6 is among the most widely used short instruments for screening psychological distress[6]. It comprises 6 items related to the following feelings during the past four weeks, such as sad, nervous, hopeless and worthless. Participants indicate their symptoms on a Likert scale ranged from 1(All of the time) to 5(None of the time). We reversed the rating of the six items to a scale from 0 to 4, and summed their scores as an indicator of psychological distress. The total score ranges from 0 to 24. The higher the scores, the higher levels the psychological distress, such as anxiety and depression. The K6 has been demonstrated good reliability and validity in Chinese populations, with Cronbach's alpha at 0.84, the 32- to 53-day interval test-retest reliability at 0.79 [15].The Cronbach alpha coefficient is 0. 88 for the female sample, 0.86 for the male sample, and 0.87 for the whole sample in the present study.

Statistical analyses

We conducted a Mokken scale analysis to examine the factor structure using the "Mokken" package for R. The package enables us to form unidimensional subscales from all items using an automated item selection algorithm (aisp). The structures of the inventory indicated by the resulting pattern and scalability of each item (expressed by Hi). We also tested the assumption of local independence and monotonicity. We concerned about whether there was any item bias between males and females, and we employed the "Lordif" package for R to detect differential item functioning (DIF) of items in the K6. Both uniform and non-uniform DIFs were detected with the logistic approach.

Results

Descriptive statistics

Responses distribution on five categories for each item in the K6 is presented in Table 1. We can see the symptoms distributed as a positive skewness. The majority of people have no symptom, while only a few

have severe symptoms. According to the cut point of 12/13, the prevalence of psychological distress among the current sample is 5.3%.

PLEASE INSERT TABLE 1 AROUND HERE

Examining factor structure

Assessment of dimensionality

Scalability of the K6 is presented in Table 2. For inter-item pairs, the inter-item scalability coefficients (Hij) range from 0.47 to 0.68. For items, the item scalability coefficients (Hi) ranged from 0.57 to 0.59. For the whole K6 scale, the scalability coefficient was 0.58(SE=0.009). All the scalability coefficients were significantly greater than the conventional lower-bound value of 0.3. The results suggested the K6 should be considered as a scale of strong strength. The internal consistency of the six items was also excellent (Cronbach's alpha =0.87).

PLEASE INSERT TABLE 2 AROUND HERE

We further explored the dimensionality for all the six items by conducting iterative automated item selection procedure (AISP). The results were presented in Table 3. We followed the recommendation of Hemker et al. (1995), and set an initial value of lower bound c from 0 to 0.75 with increment steps of 0.05. For $0 \leq c \leq 0.55$, all six items were selected to form one scale. For c=0.6, two scales emerged, including items 1-3 and items 4-6, respectively. For c=0.65, items 1 and 3 were unscalable. For c>0.7, all items were unscalable. The c value is suggested to set at 0.3 in practice, because the solution produced by the AISP is often hard to interpret when $c \geq 0.35$ (Sijtsma & van der Ark, 2017). Therefore, the results from the AISP confirmed the unidimensionality of the K6.

PLEASE INSERT TABLE 3 AROUND HERE

Assessment of local independence and monotonicity

Moreover, we examined local independence and monotonicity ensure the data were adequately fit to the Mokken scale. For local independence, no item-pair was flagged as locally dependent according to two indices (W1 and W2) calculated in the conditional association procedure [37]. That is, there is no evidence of violating local independence. For monotonicity, the results showed that no item violated the monotonicity assumption. Graphical analysis indicated that all items showed monotonical increases (see Figure 1).

PLEASE INSERT FIGURE 1 AROUND HERE

PLEASE INSERT TABLE 4 AROUND HERE

Reliability

Table 2 shows the MS method reliability-estimate. Table 1 also provides coefficients α =0.87 (Cronbach, 1951) and λ2=0.87 (Guttman, 1945). All estimates are close to .9, and thus satisfactory. The corrected item-test correlations were satisfactory for all items, ranged from 0.64 to 0.70.

Gender differences

We also conducted the same analyses on the data from the male and the female subgroups separately. A similar pattern emerged for the scalability assessment among these two samples. Therefore, the K6 assesses psychological distress in a similar way and with a similar strength both gender.

Examining measurement invariance

Following the procedure proposed by Choi et al. [32], we conducted differential item functioning (DIF) analysis under the hybrid iterative LR/IRT framework with "Lordif" package in R. Three ordinal logistic models (models 1, 2 and 3) were established for each item involving item performance, latent trait score, group membership, and the interaction between the latter two. Model 1 is a baseline model, including only the latent trait score as the predictor. Model 2 is a uniform DIF model, including the latent trait score and group membership as predictors. Model 3 is a non-uniform DIF model, including latent trait score, group membership, and their interaction as predictors. DIF detection is based on the likelihood ratio (LR) χ2 test at the α level of 0.01. A significant difference in the log-likelihood values between Model 2 and Model 1 reveals uniform DIF, while a significant difference between Model 3 and Model 2 indicates non-uniform DIF. DIF magnitude is based on McFadden's pseudo-R2, <0.13 as negligible, 0.13<R2<0.26 as moderate, >0.26 as large [33]. Under the framework, the latent trait score was estimated by default fitting Graded Response Model (GRM).

Figure 2 illustrates trait distributions of the male and the female. The male has lower mean scores than the female, but there is still a broad overlap. Table 5 presents the main results of DIF analysis. According to the LR χ2 test, Item 4 and Item 5 were marked for uniform DIF, but none was flagged for non-uniform DIF. Diagnostic plots for the two DIF items display in Figure 3 and Figure 4. Further examination of these two items revealed that for the same latent trait score, females are always rated with higher frequencies on than males. For both items, the lower-left graph shows the uniform DIF was mainly caused by the fifth category threshold value (3.31vs.2.9, 2.57 vs 2.45). However, McFadden's pseudo R^2 statistics (no more than 0.0011) indicated that the magnitude of DIF was very small for each item. Figure 5 is a graphical representation of the impact of all items and DIF items on the whole scale. The left one shows the impact of all six items, indicating negligible difference across gender. The right one shows curves for the 2 DIF items, indicating that female score a bit higher when sex group-specific parameter estimates were used.

PLEASE INSERT TABLE 5 AROUND HERE

PLEASE INSERT FIGURE 2 AROUND HERE

PLEASE INSERT FIGURE 3 AROUND HERE

PLEASE INSERT FIGURE 4 AROUND HERE

PLEASE INSERT FIGURE 5 AROUND HERE

Discussion

The present study employed a Mokken scale analysis on the K6 to evaluate its dimensionality and structure, and employed DIF analysis to examine whether the same structure existed across sex in a national representative sample of old Chinese people. The results confirmed the unidimensionality of the instrument and justified the sum score of all the six items as an indicator of psychological distress. Our study also supported the measurement invariance of the K6 between male and female populations.

The K6 was developed as a unidimensional measure for psychological distress at the beginning[5]. Later studies reported different factor solutions, one-factor models and two-factor models, with exploratory factor analysis and confirmatory factor analysis in diverse samples[13, 14, 22, 34]. The conflicting findings may result from differences in populations(e.g., emerging adults and mid-age general population) and statistical methods(e.g., principal axis factor analysis, principal component analysis). Considering the J-shape distribution of item scores in the K6, we employed a new approach, Mokken scale analysis, to address the problem in the elderly. The approach is more flexible, relying less on item score distributions and sample size[35]. Mokken scale analysis is recommended as a more appropriate method for dimensionality assessment with discrete data[36]. In addition, previous studies mainly focused on the general population, or some specific population, such as adolescents, emerging adults, but few have taped the aging population. Our findings supported the unidimensional solution, which is consistent with the original design of the K6 and most previous studies investigating the factor structure of the K6. It contributes the understanding of the sum score of all six items of the K6 as the indicator of psychological distress among aging populations.

Measurement invariance is the premise for group comparison[30]. Previous studies indicate that females always have more severe symptoms than males, but only a few studies have examined measurement invariance of the K6 across gender [21]. Drapeau et al. [11] and Mewton et al. [21] examined measurement invariance under the framework of confirmatory factor analysis. We explored measurement invariance under the LR/IRT framework and found two items were marked as with uniform DIF in term of Chi-square. For Item 4("Hopeless"), Drapeau et al. found that women had higher first three thresholds, but lower last thresholds than men. For Item 5("Everything was an effort"), they only found gender invariance only in the young age group and only at cycle 7 of the study. In the India sample, this item was separated as a second factor[6]. Mewton et al. [21] revealed that all six items have higher endorsement rates for females than males. Since the likelihood ratio test is largely influenced by sample size, DIF magnitude is also recommended to consider in detect items with DIF. In term of McFadden's pseudo R^2, the impact of the two items is negligible. Therefore, we agree with Drapeau et al. that the items in the K6 measure distress in males and females at the same degree [11]. The gender difference in the K6 scores is a

reflection of true difference in psychological distress rather than bias in reporting of the K6 items. In general, the psychological distress for females is more severe than that for males.

Some limitations should be acknowledged about the study. The present study among the few studies focused on examining psychometric properties of the K6 among a relatively large, and national representative sample of the Chinese elderly. We only focused on the general aged population here. People in different age groups endorse the items in a somewhat different way [11, 21]. Therefore, the conclusion might not apply to other age groups. In addition, the epidemiological character of psychological distress may not be the same in different cultures[12]. We should be careful before generalization of the findings to populations in other cultures. Moreover, we only investigated factor structure and gender invariance of the K6 here. Further studies can extend to other issues, such as screening efficiency in comparison with clinical diagnostic measurements.

Conclusions

We employed a Mokken scale analysis on the K6 to evaluate its dimensionality and structure, and whether the same structure existed across gender in a national representative sample of old Chinese adults. The K6 demonstrates adequate psychometric properties in the old Chinese population. It measures a unidimensional construct and holds measurement invariance across gender. The sum of all six items can be used as an indicator of non-specific psychological distress. Differences in the indicator across gender should be considered as a real difference in psychological distress between the female and the male.

Abbreviations

K6
The 6-item version of the Kessler Psychological Distress Scale
DIF
Differential item functioning
LR/IRT
Logit regression/ item response theory
EFA
Exploratory factor analysis
CFA
Confirmatory factor analysis
PCA
Principal components analysis
MSA
Mokken scale analysis
MHM
Monotone homogeneity model

AISP
Automated item selection procedure

Declarations

Ethical approval and consent to participate

All procedures performed in this study were approved by Ethics Committee of the Department of Psychology, Nanjing University. The China Family Panel Studies were approved by the Peking University Biomedical Ethics Review Committee(No. IRB00001052-14010). Informed consent was obtained from all participants included in the study.

Consent for publication

Not Applicable.

Availability of data and material

The raw data is publicly available at

https://osf.io/fq8ct/?view_only=c97f6e3bc77341d2a9bbe33f67a09c60

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Foundation of Humanities and Social Sciences, Ministry of Education of the PRC (No. 20YJA190004), and Project of Philosophy and Social Sciences from the Education Department, Jiangsu Province (No. 2018SJZDI203). They had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

LZ conceived and designed the study, and wrote the first draft of the manuscript. ZL conducted all the statistical analysis. Both authors revised the manuscript and approved the final submission.

Acknowledgements

We would like to express our thanks to the Institute of Social Science Survey of Peking University for allowing us to use the data.

References

1. Wolitzky-Taylor KB, Castriotta N, Lenze EJ, Stanley MA, Craske MG. Anxiety disorders in older adults: a comprehensive review. Depress Anxiety. 2010;27(2):190–211.
2. Fu X, Zhang K. Report on National Mental Health Development in China (2017–2018). Beijing: Social Sciences Academic Press(China); 2018.
3. Guo J, Guan L, Fang L, Liu C, Fu M, He H, Wang X. Depression among Chinese older adults: A perspective from Hukou and health inequities. J Affect Disord. 2017;223:115–20.
4. WHO. China country assessment report on ageing and health. In. Geneva. Switzerland: Author; 2016.
5. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, Walters EE, Zaslavsky AM. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med. 2002;32(6):959–76.
6. Kessler RC, Green JG, Gruber MJ, Sampson NA, Bromet E, Cuitan M, Furukawa TA, Gureje O, Hinkov H, Hu CY, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. International Journal of Methods in Psychiatric Research. 2010;19(Suppl 1):4–22.
7. Bessaha ML. Factor Structure of the Kessler Psychological Distress Scale (K6) Among Emerging Adults. Research on Social Work Practice. 2015;27(5):616–24.
8. Khan A, Chien CW, Burton NW. A new look at the construct validity of the K6 using Rasch analysis. International Journal of Methods in Psychiatric Research. 2014;23(1):1–8.
9. Tomata Y, Watanabe T, Tanji F, Zhang S, Sugawara Y, Tsuji I. The Impact of Psychological Distress on Incident Functional Disability in Elderly Japanese: The Ohsaki Cohort 2006 Study. International Journal of Environmental Research Public Health. 2018;15(11):2502.
10. Furukawa TA, Kessler RC, Slade T, Andrews G. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. Psychol Med. 2003;33(2):357–62.
11. Drapeau A, Beaulieu-Prévost D, Marchand A, Boyer R, Préville M, Kairouz S. A life-course and time perspective on the construct validity of psychological distress in women and men. Measurement invariance of the K6 across gender. BMC Med Res Methodol. 2010;10(1):68.
12. Andersen LS, Grimsrud A, Myer L, Williams DR, Stein DJ, Seedat S. The psychometric properties of the K10 and K6 scales in screening for mood and anxiety disorders in the South African Stress and Health study. International Journal of Methods in Psychiatric Research. 2011;20(4):215–23.
13. Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC. Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. International Journal of Methods in Psychiatric Research. 2010;19(Suppl 1):23–35.
14. Prochaska JJ, Sung HY, Max W, Shi Y, Ong M. Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. International Journal of Methods in Psychiatric Research. 2012;21(2):88–97.
15. Kang YK, Guo WJ, Xu H, Chen YH, Li XJ, Tan ZP, Li N, Gesang ZR, Wang YM, Liu CB, et al. The 6-item Kessler psychological distress scale to survey serious mental illness among Chinese

undergraduates: Psychometric properties and prevalence estimate. Compr Psychiatry. 2015;63:105–12.

16. Dadfar M, Vahid M, Lester D, Bahrami F. Kessler Psychological Distress Scale (K6): Psychometric testing of the Farsi form in psychiatric outpatients. Advances in Bioresearch. 2016;7:105–8.
17. Peiper N, Clayton R, Wilson R, Illback R. The performance of the K6 Scale in a large school sample. Psychol Assess. 2015;27(1):228–38.
18. Kim G, DeCoster J, Bryant AN, Ford KL. Measurement Equivalence of the K6 Scale: The Effects of Race/Ethnicity and Language. Assessment. 2016;23(6):758–68.
19. Peiper N, Lee A, Lindsay S, Drashner N, Wing J. The performance of the K6 scale in a large school sample: A follow-up study evaluating measurement invariance on the Idaho Youth Prevention Survey. Psychol Assess. 2016;28(6):775–9.
20. Vissoci JRN, Vaca SD, El-Gabri D, de Oliveira LP, Mvungi M, Mmbaga BT, Haglund M, Staton C. Cross-cultural adaptation and psychometric properties of the Kessler Scale of Psychological Distress to a traumatic brain injury population in Swahili and the Tanzanian Setting. *Health and Quality of Life Outcomes* 2018, 16(1).
21. Mewton L, Kessler RC, Slade T, Hobbs MJ, Brownhill L, Birrell L, Tonks Z, Teesson M, Newton N, Chapman C, et al. The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents. Psychol Assess. 2016;28(10):1232–42.
22. Lee S, Tsang A, Ng KL, Ma YL, Guo W, Mak A, Kwok K. Performance of the 6-item Kessler scale for measuring serious mental illness in Hong Kong. Compr Psychiatry. 2012;53(5):584–92.
23. Easton SD, Safadi NS, Wang Y, Hasson RG. The Kessler psychological distress scale: translation and validation of an Arabic version. *Health and Quality of Life Outcomes* 2017, 15(1).
24. Li Z, Zhao X, Sheng A, Wang L. Item response analysis of the Geriatric Anxiety Inventory among the elderly in China: dimensionality and differential item functioning test. BMC Geriatrics. 2019;19(1):313.
25. Mokken RJ. A Theory and Procedure of Scale Analysis. Berlin: De Gruyter; 1971.
26. Mokken RJ, Lewis C. A Nonparametric Approach to the Analysis of Dichotomous Responses. Appl Psychol Meas. 1982;6(4):417–30.
27. Hemker BT, Sijtsma K, Molenaar IW. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. Appl Psychol Meas. 1995;19(4):337–52.
28. Sijtsma K, Meijer RR, Andries van der Ark L. Mokken scale analysis as time goes by: An update for scaling practitioners. Personality Individ Differ. 2011;50(1):31–7.
29. Fushimi M, Saito S, Shimizu T, Kudo Y, Seki M, Murata K. Prevalence of psychological distress, as measured by the Kessler 6 (K6), and related factors in Japanese employees. Community Ment Health J. 2012;48(3):328–35.
30. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in crosscultural research. International Journal of Psychological Research. 2011;3(1):111–21.

31. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. J Stat Softw. 2011;39(8):1–30.
32. Wittchen HU. Screening for serious mental illness: methodological studies of the K6 screening scale. International Journal of Methods in Psychiatric Research. 2010;19(S1):1–3.
33. Zumbo BD. A Handbook on the Theory and Methods of Differential Item Functioning (DIF). Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.
34. Zhou C, Chu J, Wang T, Peng Q, He J, Zheng W, Liu D, Wang X, Ma H, Xu L. Reliability and validity of 10-item Kessler scale(K10) Chinese version in evaluation of mental health status of Chinese population. Chinese Journal of Clinical Psychology. 2008;16(6):627–9.
35. Michielsen HJ, De Vries J, Van Heck GL, De Vijver FJRV, Sijtsma K. Examination of the Dimensionality of Fatigue. European Journal of Psychological Assessment. 2004;20(1):39–48.
36. Wismeijer AAJ, Sijtsma K, Van Assen MALM, Vingerhoets AJJM. A Comparative Study of the Dimensionality of the Self-Concealment Scale Using Principal Components Analysis and Mokken Scale Analysis. J Pers Assess. 2008;90(4):323–34.

Tables

Table 1

Responses distribution on five categories

Item	0	1	2	3	4
1.Depressed	3521(54.59%)	2052(31.81%)	275(4.26%)	431(6.68%)	171(2.65%)
2.Nervous	4099(63.55%)	1782(27.63%)	193(2.99%)	296(4.59%)	80(1.24%)
3.Restless or fidgety	4007(62.12%)	1741(26.99%)	260(4.03%)	328(5.09%)	114(1.77%)
4.Hopeless	4650(72.09%)	1207(18.71%)	202(3.13%)	292(4.53%)	99(1.53%)
5.Everthing was an effort	3813(59.12%)	1599(24.79%)	278(4.31%)	545(8.45%)	215(3.33%)
6.Worthless	4790(74.26%)	1142(17.71%)	164(2.54%)	251(3.89%)	103(1.60%)

Note. *N*=6450. 1= None of the time, 2=A little of the time, 3=Some of the time, 4=Most of the time, 5=All of the time

Table 2

Descriptive statistics of the items (upper panel) and the scale (lower panel) for the K6

Item	M	SD	H_j	SE	citc
1	0.71	1.01	0.581	0.011	0.72
2	0.52	0.86	0.555	0.011	0.70
3	0.57	0.92	0.590	0.010	0.76
4	0.45	0.88	0.584	0.011	0.74
5	0.72	1.09	0.574	0.010	0.71
6	0.41	0.85	0.592	0.011	0.74
M			3.38		
SD			4.38		
H			579	0.009	
α			0.87		
λ_2			0.97		
MS			0.87		

Note. *N*=6450. H_j=item-scalability coefficient; SE=standard error of item scalability coefficient; citc= corrected item-test correlation; H=total-scalability coefficient; α=Cronbach's alpha; λ_2=Guttman's lambda-2; MS=Molenaar–Sijtsma method.

Table 3

The results of automated item selection procedure for the K6

	Item numbers			
c	Results	Scale 1	Scale 2	Unscalable
0-0.55	1: 6	1-6		
0.6	2:3,3	1-3	4-6	
0.65	2:2, 2	2, 3	4, 6	1,5
0.7-0.75	0			1-6

Table 4

Output of assessment of monotonicity

Item	#ac	#vi	#zsig	crit
1	40	0	0	0
2	40	0	0	0
3	40	0	0	0
4	40	0	0	0
5	37	0	0	0
6	40	0	0	0

Note. *N*=6450. #ac = number of active pairs that were investigated; #vi = number of violations in which the item is involved; # zsig = number of significant z-values; crit = Crit value

Table 5

Differential Item Functioning in the male and the female subgroups

Item	Uniform DIF			Non-uniform DIF	
	c_{12}^2	ΔR^2	$\Delta\beta 12$	c_{23}^2	ΔR^2
1	0.1523	0.0001	0.0024	0.5028	0.0000
2	0.0535	0.0003	0.0028	0.0406	0.0003
3	0.1448	0.0002	0.0018	0.0431	0.0003
4	0.0004	0.0011	0.0077	0.0240	0.0005
5	0.0056	0.0005	0.0062	0.4029	0.0000
6	0.9285	0.0000	0.0001	0.4288	0.0001

Figures

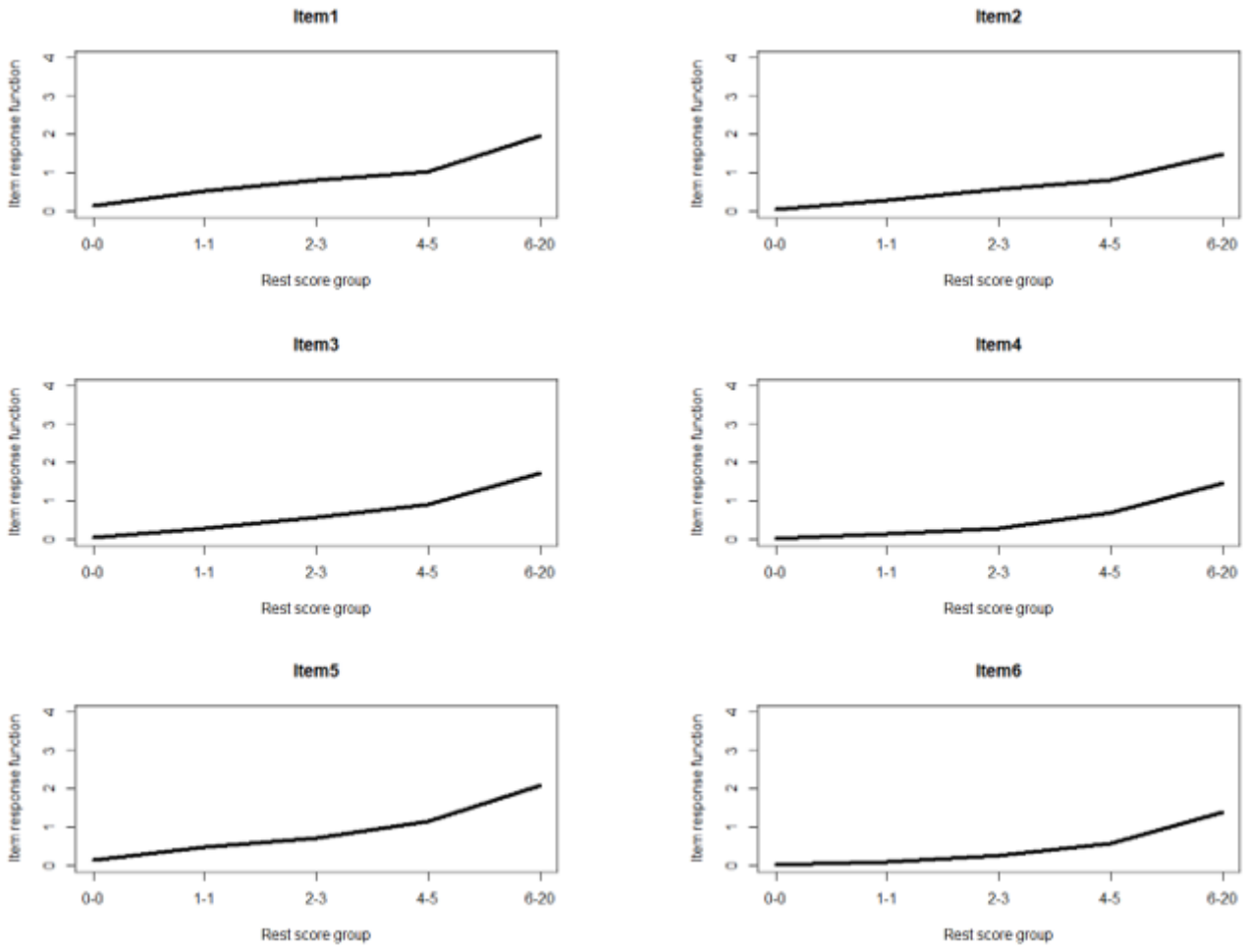

Figure 1

Monotonicity plots of the K6 items.

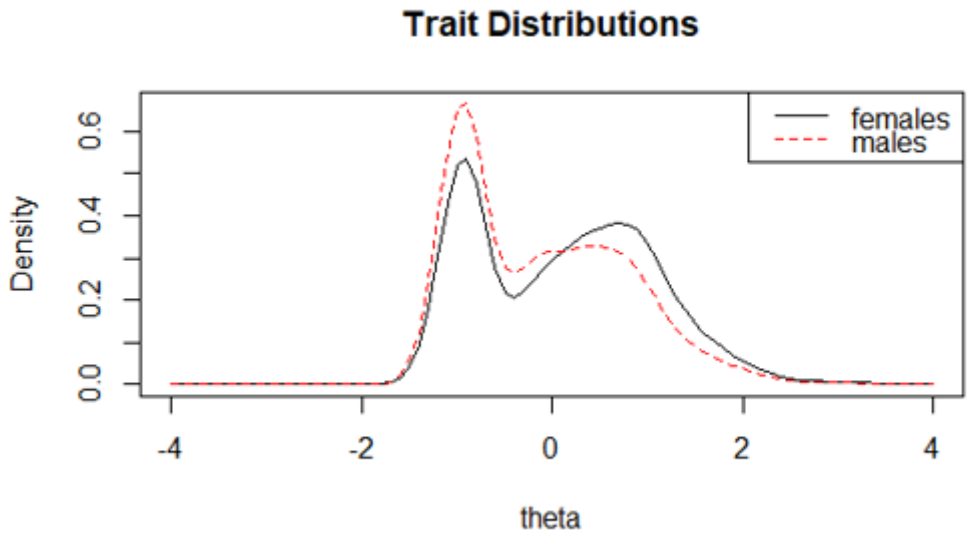

Figure 2

Trait distributions. Females (solid line) vs. Males (dashed line).

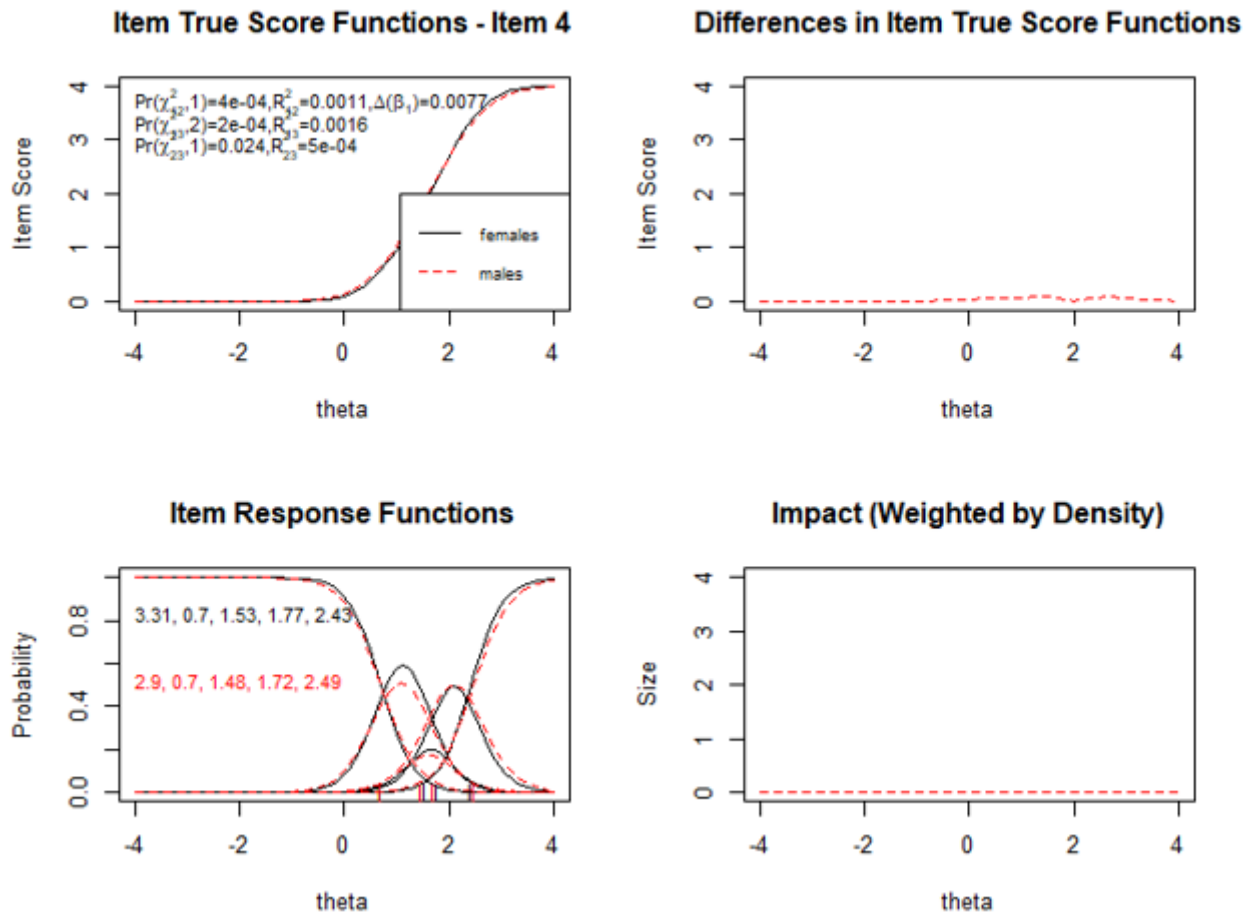

Figure 3

Graphical display of the item 4 "Hopeless". Females (solid line) vs. males (dashed line).

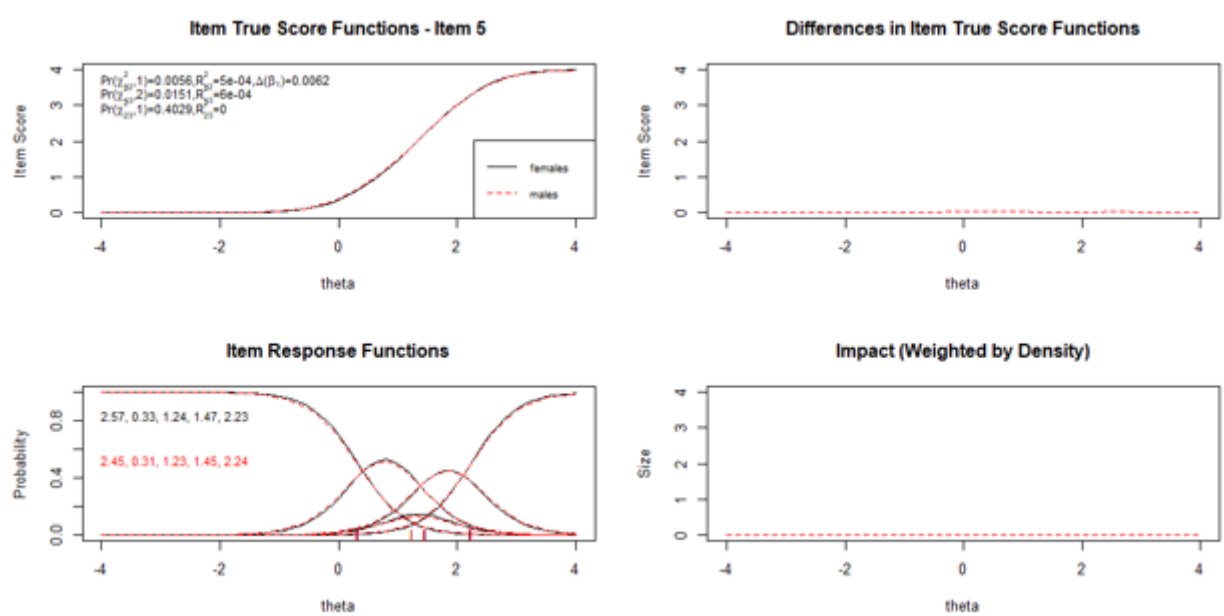

Figure 4

Graphical display of Item 5 "Everything was an effort". Females (solid line) vs. males (dashed line).

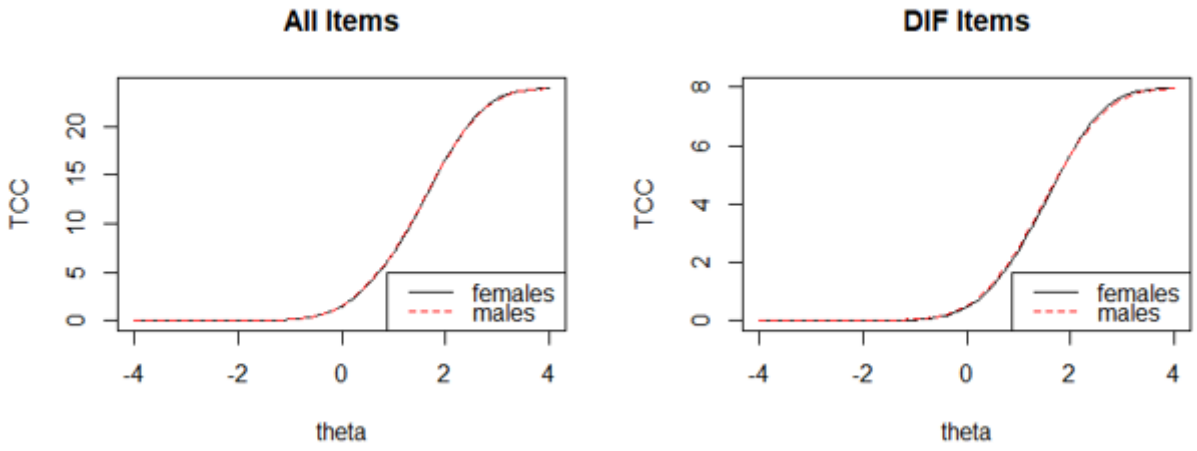

Figure 5

Impact of all items (left) and DIF items (right) on test characteristic curves. Females (solid line) vs. males (dashed line)