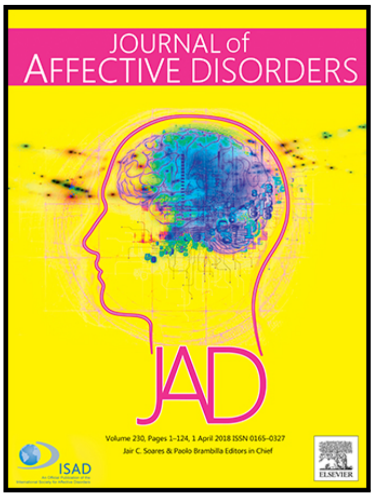
Journal Pre-proof



New avenue for the Geriatric Depression Scale: Rasch transformation enhances reliability of assessment

Alexander G. Merkin , Oleg N. Medvedev , Perminder S. Sachdev , Lynette Tippett , Rita Krishnamurthi , Susan Mahon , Nikola Kasabov , Priya Parmar , John Crawford , Zohreh G. Doborjeh , Maryam G. Doborjeh , Kristan Kang , Nicole A. Kochan , Helena Bahrami , Henry Brodaty , Valery L. Feigin

PII: S0165-0327(19)30598-1
DOI: https://doi.org/10.1016/j.jad.2019.11.100
Reference: JAD 11347

To appear in: *Journal of Affective Disorders*

Received date: 7 March 2019
Revised date: 23 October 2019
Accepted date: 21 November 2019

Please cite this article as: Alexander G. Merkin , Oleg N. Medvedev , Perminder S. Sachdev , Lynette Tippett , Rita Krishnamurthi , Susan Mahon , Nikola Kasabov , Priya Parmar , John Crawford , Zohreh G. Doborjeh , Maryam G. Doborjeh , Kristan Kang , Nicole A. Kochan , Helena Bahrami , Henry Brodaty , Valery L. Feigin , New avenue for the Geriatric Depression Scale: Rasch transformation enhances reliability of assessment, *Journal of Affective Disorders* (2019), doi: https://doi.org/10.1016/j.jad.2019.11.100

Highlights

- This article is the first to show conversion tables to transform the raw 15-item Geriatric Depression Scale (GDS-15) scores into interval-level data.
- Person Separation Index (PSI) was improved by combining locally dependent items into six super-items (PSI=0.77).
- The best Rasch model fit ($\chi2(28)=37.72$, $p=0.104$).
- Reliability and psychometric characteristics of the GDS-15 were improved and now satisfy expectations of the unidimensional Rasch model.

New avenue for the Geriatric Depression Scale: Rasch transformation enhances reliability of assessment

Authors: Alexander G. Merkin[1], Oleg N. Medvedev[2], Perminder S. Sachdev[3], Lynette Tippett[4], Rita Krishnamurthi[1], Susan Mahon[1], Nikola Kasabov[1], Priya Parmar[1], John Crawford[3], Zohreh G. Doborjeh[1], Maryam G. Doborjeh[1], Kristan Kang[3], Nicole A. Kochan[3], Helena Bahrami[1], Henry Brodaty[3], Valery L. Feigin[1]

[1] Auckland University of Technology, Auckland, New Zealand

Postal address: Faculty of Health & Environmental Sciences

National Institute for Stroke and Applied Neurosciences

AUT University

Private Bag 92006

Auckland 1142

New Zealand

amerkin@aut.ac.nz

[2] University of Waikato, School of Psychology, Hamilton, New Zealand

[3] Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia

[4] University of Auckland, Auckland, New Zealand

Abstract:

Background: Depression is a common problem in older adults. The 15-item Geriatric Depression Scale (GDS-15) is a widely used psychometric tool for measuring depression in the elderly, but its psychometric properties have not been yet rigorously investigated. The aim was to evaluate psychometric properties of the GDS-15 and improve precision of the instrument by applying Rasch analysis and deriving conversion tables for transformation of raw scores into interval level data.

Methods: The data was extracted from the prospective cohort Sydney Memory and Ageing Study of initially not demented individuals aged 70 years and older. The GDS-15 items scores of 212 participants (47.2% males) were analysed using the dichotomous Rasch model.

Results: Initially poor reliability of the GDS-15, Person Separation Index (PSI)=0.68, was improved by combining locally dependent items into seven super-items. These modifications improved reliability of the GDS-15 (PSI=0.78) and resulted in the best Rasch model fit ($\chi 2(28)=37.72$, $p=0.104$), strict unidimensionality and scale invariance across personal factors such as gender, diagnostic and language background.

Limitations: Presence of participants with cognitive impairment may be a potential limitation.

Conclusions: Reliability and psychometric characteristics of the GDS-15 were improved by minor modifications and now satisfy expectations of the unidimensional Rasch model. By using Rasch transformation tables published here psychiatrists, psychologists and researchers can transform GDS raw scores into interval-level data, which improves reliability of the GDS-15 without the need to modify its original response format. These findings increase accuracy of clinical psychometric assessments, leading to more precise diagnosis of depression in the elderly.

Keywords: Depression, Older Adults, Geriatric Depression Scale, GDS, Rasch analysis, Psychometrics.

Introduction

The number of aged people is steadily growing worldwide, therefore, mental disorders in older age have been increasingly attracting attention of specialists.(Andreas et al., 2017; Bhugra et al., 2017) Depression is one of the most common mental health issues across all age groups including older people, which represents an important public health problem associated with increased risk of morbidity and mortality, higher suicide risk, and declining cognitive and social activity.(Baune et al., 2018; Blazer, 2003; Fang et al., 2018; Haigh et al., 2018; Naushad et al., 2018; Ng, 2018; Perera et al., 2017; Read et al., 2017; Teo et al., 2018; Wang et al., 2018) Late life depression affects 1-7% of the general elderly population, rising up to 11.5% in older hospital patients and to 13.5% in those who require home healthcare.(CDC, 2017; Gonçalves-Pereira et al., 2019; Reynolds et al., 2015) Prevalence rates of depression vary from country to country, ranging from 2.4% in Japan to 35% in China.(Forlani et al., 2014; Gonçalves-Pereira et al., 2019; Haigh et al., 2018; Leles da Costa Dias et al., 2019) Such a big difference in estimations of prevalence of depression in old age can occur due to variation in diagnostic approaches in different countries, diversity of the clinical picture of depression in the elderly, and clinical overlap of depression symptoms with other mental disorders occurring in older age, primarily dementia and delirium.(Grayson and Thomas, 2013; Haigh et al., 2018) All that may lead to misdiagnosis and subsequent inappropriate treatment.(Grayson and Thomas, 2013; Haigh et al., 2018) Therefore, differential diagnosis of depression in old age represents a significant challenge for clinicians.

There is a number of psychometric instruments used for screening for depression.(Ballard et al., 2018; El-Den et al., 2018; Taouk et al., 2017; Tsoi et al., 2017) The Geriatric Depression Scale (GDS) is one of the most widely used psychometric tools designed specifically for measuring depression in the elderly.(He et al., 2018; Sjöberg et al., 2017; Tsoi et al., 2017; Yesavage et al., 1982) The GDS focuses on the psychological aspects of depression rather than somatic symptoms to avoid interference from comorbid physical conditions that are common in old age.(Blumenthal, 1975; Chiesi et al., 2017b;

Yesavage et al., 1982) The 15-item Geriatric Depression Scale (GDS-15) is a shorter version of the 30-item version intended for use as a screening tool for geriatric depression rather than for diagnostic classification, and requires shorter time and effort to complete, which is an important advantage if working with older populations.(Sheikh and Yesavage, 1986) There have been several reviews of the diagnostic accuracy of the GDS-15,(Chiesi et al.; Chiesi et al., 2018; Chiesi et al., 2017b; Guerin et al., 2018; Niu et al., 2018; Sheikh and Yesavage, 1986) indicating a pooled sensitivity of 0.89 and specificity 0.77 at the recommended cut-off score of 5 and greater, and reported of the widespread use of the GDS-15 in both practice and research across different linguistic and clinical settings.(Chiesi et al., 2017b; Pocklington et al., 2016) The scale has demonstrated satisfactory reliability and converging validity using more traditional testing methods: internal consistency (Cronbach's alpha) range from 0.76 to 0.83.(Iglesias, 2004; Van Marwijk et al., 1995) However, a traditional reliability estimate such as Cronbach's alpha only evaluates consistency across scale items in assessing the overall construct of depression. It cannot estimate how accurate the scale is in discriminating between people with different depression levels.

Rasch analysis is one of the most precise examinations of item parameters and scale properties, and is able to accurately estimate scale ability to differentiate between individual levels of depression.(Christensen et al., 2013; Rash, 1960) Importantly, Rasch analysis precisely estimates relative input of each symptom/item into the whole diagnosis based on a particular diagnostic scale, which refers to item difficulty parameter. For instance, the GDS reversed scored item 13 "Do you feel full of energy?" has a low contribution to the overall depression score and is considered an easy item to endorse. In contrast the item 11 "Do you think it is wonderful to be alive now?", that is reversed scored with 'No' corresponding to higher depression, is difficult to endorse for an average person because it reflects higher depression levels as indicated by an earlier Rasch analysis.(Chiang et al., 2009) In everyday clinical practice these items are rendered as having the same value and the total score is computed by adding individual items scores, which results in diagnostic bias that can have serious implications for a patient. Rasch analysis enables such assessment bias to be eliminated by accounting for true contributions of each item to the overall depression scores. The Rasch model is identical to the one parameter Item Response Theory (IRT) model.(Hobart and Cano, 2009) It is a probabilistic, unidimensional measurement model stating that response to a specific item is a function of only two parameters: personal ability and item difficulty, which are estimated independently using the same metric.(Rash, 1960) When data fit the Rasch model, transformation of the raw scale scores into interval-level data becomes possible, increasing accuracy of measurement. This is the main advantage of the Rasch model over conventional statistical methods. Rasch analysis is beneficial for investigating and improving psychometric properties of ordinal scales, as it enhances accuracy of measurement and provides better covering of a range of sample abilities (e.g. depression levels) that have been established both theoretically and empirically.(Christensen et al., 2013) Both the IRT and Rasch models has been whidely applied evaluate and improve psychometric measurement in psychiatry, psychology, and rehabilitation medicine.(Allega et al., 2018; Ayis et al., 2018; Brunton et al., 2018; Hobart and Cano, 2009; Lundgren and Tennant, 2011; Svicher et al., 2019)

Tang et al (2005) and Chiang et al (2009) applied Rasch analysis to the GDS-15, and reported adequate functioning of all 15 items.(Chiang et al., 2009; Tang et al., 2005) It was advised the GDS-15 may be less effective as a screening tool but could be better used to detect a change in moderate levels of depression.(Chiang et al., 2009) However, there were some limitations of these studies. First, one of them was conducted using the Chinese scale version on a sample that was drawn from Chinese patients with pneumoconiosis and may not be applicable to non-Chinese populations or patients with other forms of medical disease,(Tang et al., 2005) whereas the other study was based on a relatively small sample of participants (n=177).(Chiang et al., 2009) Moreover, Person Separation Index (PSI), which estimates scale ability to discriminate between individual depression levels, was not reported although it is an essential measure of reliability in the Rasch analysis that represents a reliable alternative to sensitivity analysis. There were also no converging tables generated to convert raw scores into interval level data, which is the main advantage of the Rasch analysis in line with recently published recommendation.(Leung et al., 2014) Therefore, the aim of this study was to apply Rasch analysis to evaluate psychometric properties of the GDS-15 and produce conversion tables for transformation of raw scores into interval level data to enhance accuracy of measurement.

Method

Participants

The study was approved by the Auckland University of Technology Ethics Committee (AUTEC) (17/394). The sample initially consisted of 1037 participants extracted from the prospective community-based cohort Sydney Memory and Ageing Study (MAS study)(Sachdev et al., 2010) of non-demented individuals aged 70 years and older followed up bi-annually. The MAS study was initiated in 2005 to examine the clinical characteristics and prevalence of mild cognitive impairment (MCI) and related syndromes in older Australians, and to determine the rate of change in cognitive function over time.(Sachdev et al., 2010) The data was taken from the Wave 1 of the MAS study, which represents the first assessment in turn, without participants diagnosed with dementia. Figure 1 shows the stages of sample extraction. We removed 4 participants from the analysis for the GDS-15 because they did not complete the GDS assessment. In order to have participants with different depression levels equally represented in the sample, we extracted participants with the total GDS-15 score equal or more than 5 as they represented the minority of this non-clinical sample (106 individuals, 15%). We used 5 as a cut-off score for depression in accordance with recommendations of Sheikh and Yesavage (1986).(Sheikh and Yesavage, 1986) We then randomly selected an equal number (106) of participants from the remaining 927 participants, as shown in Figure 1.

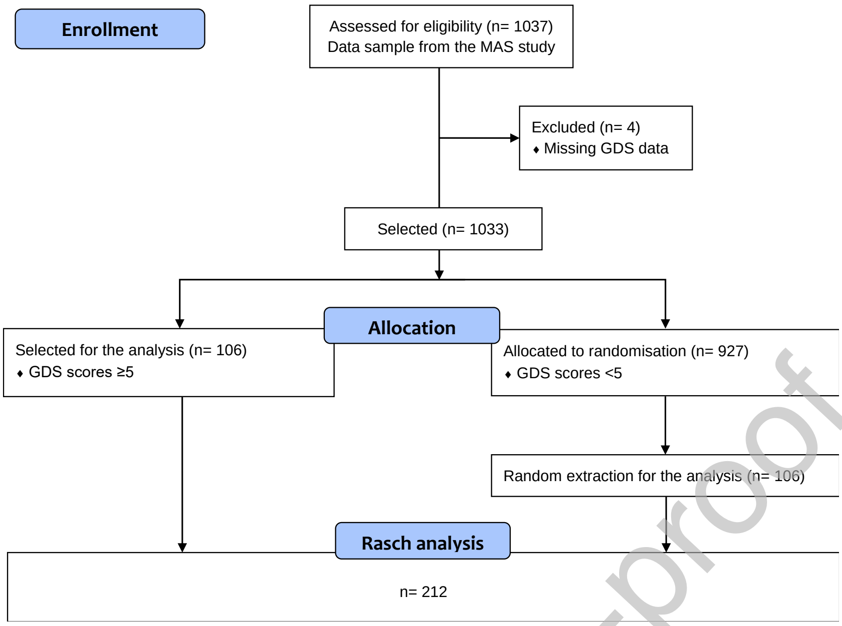


Figure 1. Consort flow diagram of the study population extracted for the Rasch analysis of the GDS-15.

These two samples were combined into one sample (n=212), which satisfies requirements of 10 person per item for dichotomous Rasch model indicated by the similar study.(Chiang et al., 2009) The sub-sample included participants with the age range from 71 to 90 years (mean=79.1 years, SD=5.0), 100 (47.2%) of them were males. Out of 212 participants 42 (19.8%) were from non-English speaking background, but fluent in English. One hundred participants (47.2%) were cognitively healthy and 72 (34.0%) diagnosed with mild cognitive impairment (MCI), and 40 (18.9%) participants were unclassified due to missing diagnostic data.(Sachdev et al., 2010) This fact could not intervene in the results considering limited influence of the cognitive abilities on the indicative of depression with GDS. The GDS total scores were distributed close to normal with maximum values of 0.75 for skewness and 0.44 for kurtosis falling within the conservative range of ±1. The mean depression score was 4.37 (SD=3.12) ranging from 0 to 15. There were no significant differences in the distribution of sex and age between the full sample and the extracted sub sample (both p >0.05).

GDS-15 scale used in the MAS study

The GDS-15 contains 15 dichotomous items in which participants are asked to respond by answering "yes" or "no" about how they feel (Table 1). Ten out of 15 items indicate the presence of depression when answered positively, while the other items including one, five, seven, eleven and thirteen indicate depression when answered negatively. Scores from 0 to 4 indicate absence of depression; 5 to 8 correspond to mild depression; 9 to 11 correspond to moderate depression; and 12 to 15 – to severe depression.(Sheikh and Yesavage, 1986) Cronbach's alpha with the current dataset (n=212) was 0.79, which is consistent with previous reports.(Iglesias, 2004; Van Marwijk et al., 1995)

Table 1. The 15-item Geriatric Depression Scale.

1. Are you basically satisfied with your life?
2. Have you dropped many of your activities and interests?
3. Do you feel that your life is empty?
4. Do you often get bored?
5. Are you in good spirits most of the time?
6. Are you afraid that something bad is going to happen to you?
7. Do you feel happy most of the time?
8. Do you often feel helpless?
9. Do you prefer to stay at home at night, rather than going out and doing new things?
10. Do you feel you have more problems with memory than most?
11. Do you think it is wonderful to be alive now?
12. Do you feel pretty worthless the way you are now?

13. Do you feel full of energy?

14. Do you feel that your situation is hopeless?

15. Do you think that most people are better off than you are?

MAS used the GDS-15 with item 9 as described in Brink (1982, here item 12)(Brink, 1982) instead of the more common one. In order to determine whether this would affect classification of depression, a statistical analysis was carried out on two data sets from the same sample, one with the GDS-15 containing the version used in MAS, and the other with the more common version. Using the same cut-point for depression (score ≥6) the results showed a high level of agreement between the two classifications (kappa = .931) with a total of only 4 false positives, and no false negatives, out of a sample of 354 used in MAS study to validate this item.(Sachdev et al., 2010) All items are presented in Table 1.

Data Analyses

Descriptive statistics were computed using IBM SPSS v.25 and Rasch analysis was conducted using Rasch model software package RUMM2030. Missing data in all GDS items comprised 0.9% and were completely at random. Rasch analysis was conducted in iterative fashion following the sequential main steps (Tennant and Conaghan, 2007):

1. A test for overall data fit to the Rasch model.
2. Identifying items with poorest fit to the Rasch model and local dependency between items.
3. Analysis of Differential Item Functioning (DIF) for personal factors.
4. Scale modifications: deletion of misfitting items, resolving local dependency/creating super-items.
4. Re-testing individual item fits and the overall fit to the Rasch Model.
5. Unidimensionality test.
6. Distribution analysis of the participant-item thresholds.

7. Conversion of raw scores into interval measure based on person estimates.

The initial analysis started with evaluation of the overall data fit and individual item's fit to the Rasch and included examination of residual correlations between items because it may affect fit to the model. Rasch model fit was evaluated using the following criteria: the overall Rasch model fit requires non-significant item-trait interaction indexed by chi square (p>0.05, Bonferroni adjusted). Overall fit residuals for both item and person should have mean close to 0.00 and a standard deviation close to 1.00. Individual items should have fit residuals between -2.50 and +2.50, and no items should work significantly differently for different groups (e.g. male vs females), which refers to Differential Item Functioning (DIF) tested using ANOVA between relevant sample groups. Reliability in Rasch analysis is evaluated using PSI, which is numerically close to Cronbach's alpha.

Rasch analysis involves an iterative process that begins with testing the overall and individual item fit to the model, local dependency between items and DIF. Usually, items displaying poor-fit or DIF are deleted and locally dependent items can be combined into super-items, and the overall model fit is then re-tested.(Tennant and Conaghan, 2007) Residual correlations that exceed the cut-off point of 0.2 indicating local dependency between relevant items.(Christensen et al., 2017) Rasch analysis is completed once both the overall model and individual item fit to the model are acceptable and evidence for unidimensionality of the scale is obtained. In this analysis we considered deletion of an item as the last resort because all items of the GDS-15 short version represent significant depression symptoms and further reduction of the scale may affect validity of the construct. This approach is also supported by findings of Tang et al (2005) and Chiang et al (2009) who reported adequate functioning of all 15 items.(Chiang et al., 2009; Tang et al., 2005)

In Rasch analysis, dimensionality is normally tested by examining the first principal component of the residuals after the main latent trait (e.g. depression) is removed(Smith Jr, 2002). An independent t-test is then conducted to compare person-locations for the two groups of items: one group with the highest positive and the other group with lowest negative loadings on the first principal component of the residuals. In the current study unidimensionality was tested using the standardized unweighted item fit residuals(Smith Jr, 2002). Unidimensionality is evident if the percentage of significant *t*-test comparisons is below 5% or if the lower bound of the binominal confidence interval computed for the number of significant *t*-tests overlaps 5%. When a satisfactory Rasch model fit is evident, the person-item thresholds distribution is evaluated to determine how well item thresholds cover depression levels of the sample. Finally, the conversion tables are generated based on person estimates of the Rasch model that permit transformation of raw scores into interval-level data to increase accuracy of assessment.

Role of the funding source

The funder of the study had no role in study design, data collection, analysis, and interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Initial analysis of the full scale of 15 GDS items indicated an overall poor fit to the Rasch model with significant chi square for interaction between items and the latent trait ($\chi^2(75)=211.78$, $p<0.0001$) and low reliability PSI=0.68. Table 2 includes the Rasch model fit statistics for the initial analysis on the individual items and indicates five items with significant model misfit. The strongest misfit was associated with item 9 (Do you prefer to stay home?), displaying fit residual of 4.37. The item 7 (Do you feel happy most of the time?) also exhibited significant negative fit residual at the level of 3.09. The

Rasch model fit may be affected by local dependency between items. Therefore, the residual correlation matrix was examined and indicated residual correlations between most of the items that exceeded the cut-off point of 0.2 indicating local dependency.(Christensen et al., 2017)

Table 2. Rasch model GDS-15 items fit statistics for the initial analysis.

Item	Location	SE	FitResid	$p_{FitResid}$	ChiSq	p_{ChiSq}
1	0.86	0.20	-1.96	0.004	15.33	0.000
2	-0.81	0.16	2.74*	0.004	15.35	0.003
3	0.68	0.20	-2.97*	0.067	8.78	0.007
4	0.29	0.18	-0.96	0.041	9.97	0.031
5	0.58	0.19	-1.97	0.118	7.37	0.064
6	0.09	0.18	-0.91	0.137	6.98	0.140
7	-0.15	0.17	-3.09*	0.002	17.13	0.000
8	0.30	0.18	-1.93	0.295	4.93	0.275
9	-2.26	0.18	4.37*	0.000	67.61	0.000
10	0.29	0.18	0.49	0.354	4.41	0.524
11	0.30	0.18	-2.00	0.050	9.51	0.015
12	0.11	0.18	-2.10	0.139	6.94	0.111
13	-2.27	0.18	2.58*	0.000	24.78	0.000
14	1.21	0.23	-2.39	0.012	12.81	0.000
15	0.78	0.20	-1.92	0.086	8.15	0.018

Note: *Significant misfit to the Rasch model (p<0.01).

Instead of deleting items, local dependency issue can be resolved by combining locally dependent items into super-items that sum scores of two or more dependent items together into a single variable, which reduces measurement error.(Medvedev et al., 2018) Therefore, locally dependent items were combined into super-items as follow: Super-item 1 (Items 7+9+10); Super-item 2 (5+8); Super-item 3

(1+2); Super-item 4 (11+15); 5 (3+13); 6 (6+12); 7 (4+14). After creating super-items the best model fit was achieved with non-significant chi square for item-trait interaction (χ^2(28)=37.72, *p*=0.104). This modification resulted in substantial improvement of reliability (PSI=0.78), strict unidimensionality (2.8% of significant *t*-tests) and no differential item functioning by age, sex and diagnosis (e.g. MCI). Table 3 displays Rasch model fit statistics for individual super-items of the modified GDS-15. At this stage there were no more misfitting items and no local dependency was evident between super-items.

Table 3. The overall Rasch model fit statistics for the modified GDS-15 super-items.

Item	Location	SE	Fit Residual	$p_{FitResid}$	Chi Square	p_{ChiSq}
Super-item 1	-1.01	0.14	-0.74	0.190	6.12	0.092
Super-item 2	0.69	0.15	-1.64	0.081	8.32	0.017
Super-item 3	-0.04	0.15	0.73	0.218	5.76	0.191
Super-item 4	0.60	0.15	-1.20	0.680	2.30	0.738
Super-item 5	-1.20	0.15	0.24	0.854	1.34	0.843
Super-item 6	0.19	0.14	-0.65	0.048	9.56	0.036
Super-item 7	0.76	0.15	-1.00	0.365	4.31	0.349

Note: Super-item 1 (Items 7+9+10); 2 (5+8); 3 (1+2); 4 (11+15); 5 (3+13); 6 (6+12); 7 (4+14)

Figure 2 shows person-item threshold distribution of the modified GDS-15 by creating super-items. The scale's thresholds satisfactorily cover depression levels of the sample without any significant ceiling or floor effects. Sample mean of the sample is below items mean indicating that overall low depression levels in the sample. This indicates good potential of the modified GDS-15 to assess populations with high levels of depression.

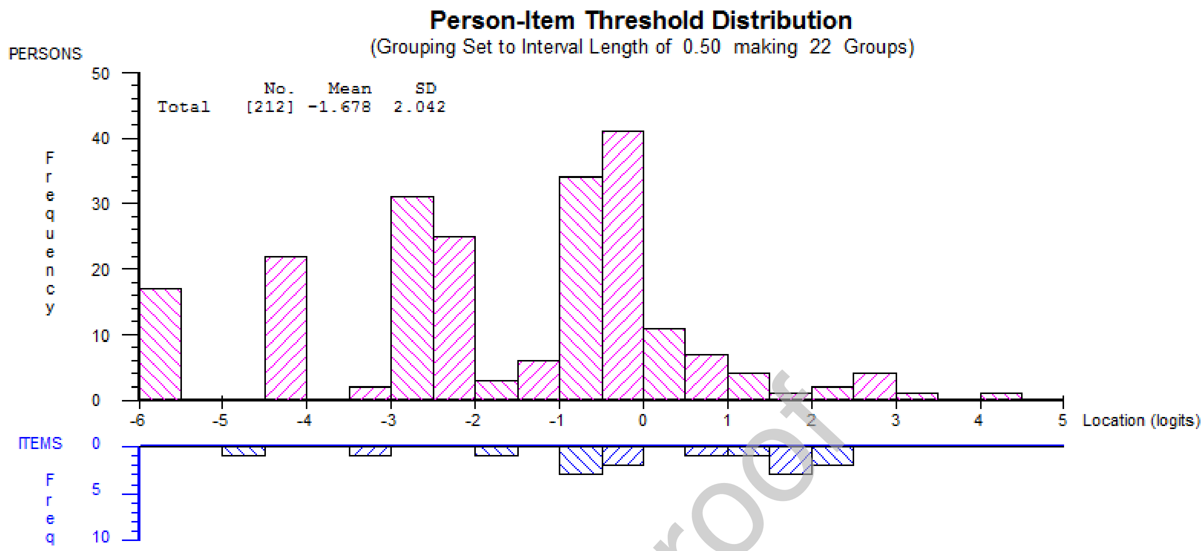


Figure 2. Person-item threshold distribution of the final modified GDS-15.

The satisfactory Rasch model fit of the GDS-15 permitted transformation of raw scale scores into interval scale by estimating genuine interval-level scores for each ordinal raw score. Developed Rasch interval transformation table permits conversion of the raw ordinal test scores into interval-level data (see Table 4).

Table 4. Rasch interval transformation table.

Ordinal Scale	Interval Scale		±95% Confidence Intervals
Raw Scores	Logits	Scale	[±95% CI]
0	-5.94	0.00	3.76
1	-4.10	2.74	2.41
2	-2.86	4.60	1.94
3	-2.02	5.85	1.65

4	-1.41	6.77	1.47
5	**-0.91**	**7.52**	1.37
6	-0.46	8.19	1.31
7	-0.04	8.81	1.27
8	0.36	9.42	1.25
9	0.76	10.01	1.25
10	1.16	10.61	1.27
11	1.57	11.22	1.31
12	2.02	11.89	1.39
13	2.53	12.66	1.55
14	3.21	13.67	1.88
15	4.10	15.00	2.63

Discussion

The 15-item Geriatric Depression Scale (GDS-15) is one of the most widely used psychometric instruments in the elderly. The present study significantly enhanced measurement properties of the GDS-15 by applying Rasch analysis. The current work is novel, because it developed conversion tables to transform GDS-15 raw binary scores into interval level data, which significantly increases accuracy and sustainability of depression screening. Theoretically, meeting expectations of the Rasch model achieved for GDS-15 in this study means that the scale complies with principles of the fundamental measurement common to physical phenomena such as invariance across population groups, unidimensionality and the same measurement unit across scale continuum. (Rash, 1960) We used modern Rasch methodology that involved combining locally dependent items into super-items to improve reliability and internal validity without the need to delete scale items.(Medvedev et al., 2018) After these minor modifications the GDS-15 demonstrated excellent coverage of depression levels in the current sample with the capacity to reliably assess higher depression levels in clinical populations (Figure 2). Using super-items is beneficial because it reduces measurement error associated with individual items as every item may measure some other psychological aspects to a different degree. For instance, item 15 (Do you think that most people are better off than you are?) besides depression also measures self-esteem, which affects internal scale validity. When such items are combined based on residual correlations, individual item errors cancel each other resulting in improvement of both reliability and validity, which is not achievable by simply adding the raw scores following current measurement practice.

Combination of some items to super-items is also plausible based on a single nature of a psychological facet they describe. For example, the super-item 2, which combines the reversed coded item 5 (Are you in good spirits most of the time?) and item 8 (Do you often feel helpless?) may be described as a feeling of devastation. The reversed coded item 1 (Are you basically satisfied with your life?) may be outlined as an outcome of the item 2 (Have you dropped many of your activities and interests?), therefore combining those two items into the super-item 3 appears meaningful. The item 3 (Do you feel that your life is empty?) and reversed coded item 13 (Do you feel full of energy?) represent liveliness and resourcefulness and pooled into the super-item 5. The super-item 7 is a result of integration of the item 4 (Do you often get bored?) and item 14 (Do you feel that your situation is hopeless?) because they both may be associated with loosing meaning of life. At the same time, there are some combinations of items into super-items that represent highly variable system of meanings. For instance, the super-item 4 is a result of putting together the reversed coded item 11 (Do you think it is wonderful to be alive now?) and 15 (Do you think that most people are better off than you are?). If considering that the item 11 is negatively worded, it is plausible that somebody with low self-esteem (item 15) may not feel happy with their life. Please note that most of the created super-items combine normal depression items with reverse coded items (measuring opposite of depression) and hence eliminate a method effect associated with measurement error due to item wording (negative vs positive).

Transformation of raw scores presented in this article account for all these biases evidenced by the best fit to the fundamental measurement Rasch model achieved using super-items method.(Medvedev et al., 2018) Moreover, transformation of raw scores into interval scale presented here does not require changes in test administration, it just exploits a different scoring algorithm of the total score. This transformation can be conducted by finding individual item scores (original raw scores) in the column 1 of the conversion table provided here (Table 4), and finding the corresponding interval-level score (transformed score) in the column 3 of the conversion table . Interval-level scores are presented using the original scale range, which is an additional advantage because it permits comparisons of raw and interval scores. The results of the study can be used by different medical professionals (psychiatrists, neurologists, psychologists, nurses, caregivers) and researchers who apply original 15-items GDS version. This conversion is very simple and can be conducted by specialists not familiar with statistical analysis. The modified GDS-15 scale has a good potential to assess populations with different levels of depression from low to high equally effective.

Interval-level data transformed with Rasch analysis is also suitable for parametric statistical tests such as ANOVA without violating their fundamental assumptions. It simplifies statistical data analyses by allowing valid comparison with other interval measures and increases accuracy of assessment by reducing measurement error due to item misfit in comparison with ordinal scores. Using this Rasch conversion tables also permits valid comparisons of GDS-15 interval data with electrophysiological and neuroimaging data such as EEG, (f)MRI as well as biomarkers and genetics data, which especially important for researchers who use GDS-15 as a psychometric tool.

Limitations

A possible limitation of our study relates to the actual ability of participants with cognitive impairment to provide accurate responses to GDS items as per work of Chiang et al.(Chiang et al., 2009)

However, the findings by Chiesi et al (2018) referred in the manuscript that support usage of the GDS-15 for this group of people.(Chiesi et al., 2018)

In summary, the study provides modified GDS-15 as a reliable and valid screening tool for assessment of depression in older adults. Such an improvement of the GDS-15 scale is facilitated by Rasch transformation of raw scores into interval-level data, which contribute both to clinical and scientific purposes and permit wide application of the scale in clinical and research facilities. It will enhance accuracy of assessment for screening, early detection and treatment of depression and dementia in the elderly.

Acknowledgments

The data from the Memory and Ageing Study was provided by the MAS Management Committee. This research was supported by a research grant from the Brain Research New Zealand Centre of Research Excellence.

Declaration of interests

The authors declare that there is no conflict of interest.

Author Details

- Dr Alexander G. Merkin, Research Fellow, Auckland University of Technology, Auckland, New Zealand.

Postal address: Faculty of Health & Environmental Sciences

National Institute for Stroke and Applied Neurosciences

AUT University

Private Bag 92006

Auckland 1142

New Zealand

- Dr Oleg N. Medvedev, Auckland University of Technology, Auckland, New Zealand.
- Scientia Professor of Neuropsychiatry Perminder S. Sachdev, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia.
- Associate Professor Lynette Tippett, University of Auckland, Auckland, New Zealand.
- Dr Rita Krishnamurthi, Auckland University of Technology, Auckland, New Zealand.
- Dr Susan Mahon, Auckland University of Technology, Auckland, New Zealand.
- Professor Nikola Kasabov, Auckland University of Technology, Auckland, New Zealand.
- Dr Priya Parmar, Auckland University of Technology, Auckland, New Zealand.
- Dr John Crawford, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia.
- PhD candidate Zohreh G. Doborjeh, Auckland University of Technology, Auckland, New Zealand.
- Maryam G. Doborjeh, Auckland University of Technology, Auckland, New Zealand.
- Dr Kristan Kang, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia.
- Dr Nicole A. Kochan, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia.
- PhD candidate Helena Bahrami, Auckland University of Technology, Auckland, New Zealand.

- Scientia Professor Henry Brodaty, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, Australia.
- Professor Valery L. Feigin, Auckland University of Technology, Auckland, New Zealand.

Author Contributions

- Alexander G. Merkin conducted literature search, developed study design, conducted data analysis and interpretation, authored or reviewed drafts of the paper, prepared figures and/or tables, approved the final draft, submitted the manuscript.
- Oleg N. Medvedev conceived and designed the study, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Perminder S. Sachdev completed data collection, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Lynette Tippett conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Rita Krishnamurthi conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Susan Mahon conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Nikola Kasabov conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Priya Parmar developed study design, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.

- John Crawford completed data collection, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Zohreh G. Doborjeh conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Maryam G. Doborjeh conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Kristan Kang completed data collection, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Nicole A. Kochan completed data collection, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Helena Bahrami conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Henry Brodaty completed data collection, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.
- Valery L. Feigin developed study design, conducted data analysis and interpretation, authored or reviewed drafts of the paper, approved the final draft.

References

Allega, O.R., Leng, X., Vaccarino, A., Skelly, M., Lanzini, M., Hidalgo, M.P., Soares, C.N., Kennedy, S.H., Frey, B.N., 2018. Performance of the biological rhythms interview for assessment in neuropsychiatry: An item response theory and actigraphy analysis. Journal of Affective Disorders 225, 54-63.
Andreas, S., Schulz, H., Volkert, J., Dehoust, M., Sehner, S., Suling, A., Ausín, B., Canuto, A., Crawford, M., Da Ronch, C., 2017. Prevalence of mental disorders in elderly people: the European MentDis_ICF65+ study. The British Journal of Psychiatry 210, 125-131.
Ayis, S.A., Ayerbe, L., Ashworth, M., Wolfe, C.D., 2018. Evaluation of the Hospital Anxiety and Depression Scale (HADS) in screening stroke patients for symptoms: Item Response Theory (IRT) analysis. Journal of Affective Disorders 228, 33-40.

Ballard, E.D., Yarrington, J.S., Farmer, C.A., Lener, M.S., Kadriu, B., Lally, N., Williams, D., Machado-Vieira, R., Niciu, M.J., Park, L., Zarate, C.A., 2018. Parsing the heterogeneity of depression: An exploratory factor analysis across commonly used depression rating scales. Journal of Affective Disorders 231, 51-57.
Baune, B.T., Malhi, G.S., Morris, G., Outhred, T., Hamilton, A., Das, P., Bassett, D., Berk, M., Boyce, P., Lyndon, B., Mulder, R., Parker, G., Singh, A.B., 2018. Cognition in depression: Can we THINC-it better? Journal of Affective Disorders 225, 559-562.
Bhugra, D., Tasman, A., Pathare, S., Priebe, S., Smith, S., Torous, J., Arbuckle, M.R., Langford, A., Alarcón, R.D., Chiu, H.F.K., 2017. The WPA-lancet psychiatry commission on the future of psychiatry. The Lancet Psychiatry 4, 775-818.
Blazer, D.G., 2003. Depression in late life: review and commentary. The journals of gerontology. Series A, Biological sciences and medical sciences 58, 249-265.
Blumenthal, M.D., 1975. Measuring depressive symptomatology in a general population. Archives of General Psychiatry 32, 971-978.
Brink, T., 1982. Geriatric depression and hypochondriasis: Incidence, interaction, assessment and treatment. Psychotherapy: Theory, Research & Practice 19, 506.
Brunton, R.J., Dryer, R., Krägeloh, C., Saliba, A., Kohlhoff, J., Medvedev, O., 2018. The pregnancy-related anxiety scale: A validity examination using Rasch analysis. Journal of Affective Disorders 236, 127-135.
CDC, 2017. Depression is Not a Normal Part of Growing Older.
Chiang, K.S., Green, K.E., Cox, E.O., 2009. Rasch analysis of the geriatric depression scale–short form. The Gerontologist 49, 262-275.
Chiesi, F., Primi, C., Pigliautile, M., Baroni, M., Ercolani, S., Boccardi, V., Ruggiero, C., Mecocci, P., 2017a. Is the 15-item Geriatric Depression Scale a Fair Screening Tool? A Differential Item Functioning Analysis Across Gender and Age. Psychological Reports 0, 0033294117745561.
Chiesi, F., Primi, C., Pigliautile, M., Baroni, M., Ercolani, S., Paolacci, L., Boccardi, V., Mecocci, P., 2018. Does the 15-item Geriatric Depression Scale function differently in old people with different levels of cognitive functioning? Journal of affective disorders 227, 471-476.
Chiesi, F., Primi, C., Pigliautile, M., Ercolani, S., della Staffa, M.C., Longo, A., Boccardi, V., Mecocci, P., 2017b. The local reliability of the 15-item version of the Geriatric Depression Scale: An item response theory (IRT) study. Journal of psychosomatic research 96, 84-88.
Christensen, K.B., Kreiner, S., Mesbah, M., 2013. Rasch models in health. John Wiley & Sons.
Christensen, K.B., Makransky, G., Horton, M., 2017. Critical values for Yen's Q 3: identification of local dependence in the Rasch model using residual correlations. Applied psychological measurement 41, 178-194.
El-Den, S., Chen, T.F., Gan, Y.-L., Wong, E., O'Reilly, C.L., 2018. The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. Journal of Affective Disorders 225, 503-522.
Fang, X., Zhang, C., Wu, Z., Peng, D., Xia, W., Xu, J., Wang, C., Cui, L., Huang, J., Fang, Y., 2018. Prevalence, risk factors and clinical characteristics of suicidal ideation in Chinese patients with depression. Journal of Affective Disorders 235, 135-141.
Forlani, C., Morri, M., Ferrari, B., Dalmonte, E., Menchetti, M., De Ronchi, D., Atti, A.R., 2014. Prevalence and gender differences in late-life depression: a population-based study. The American Journal of Geriatric Psychiatry 22, 370-380.
Gonçalves-Pereira, M., Prina, A.M., Cardoso, A.M., da Silva, J.A., Prince, M., Xavier, M., 2019. The prevalence of late-life depression in a Portuguese community sample: A 10/66 Dementia Research Group study. Journal of Affective Disorders 246, 674-681.

Grayson, L., Thomas, A., 2013. A systematic review comparing clinical features in early age at onset and late age at onset late-life depression. Journal of Affective Disorders 150, 161-170.
Guerin, J.M., Copersino, M.L., Schretlen, D.J., 2018. Clinical utility of the 15-item geriatric depression scale (GDS-15) for use with young and middle-aged adults. Journal of Affective Disorders 241, 59-62.
Haigh, E.A.P., Bogucki, O.E., Sigmon, S.T., Blazer, D.G., 2018. Depression Among Older Adults: A 20-Year Update on Five Common Myths and Misconceptions. The American Journal of Geriatric Psychiatry 26, 107-122.
He, J., Zhong, X., Yao, S., 2018. Factor structure of the Geriatric Depression Scale and measurement invariance across gender among Chinese elders. Journal of Affective Disorders 238, 136-141.
Hobart, J., Cano, S., 2009. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Prepress Projects.
Iglesias, G.H., 2004. Geriatric Depression Scale short form and Zung Self-rating Depression Scale: A study of homebound elders. Clinical gerontologist 27, 55-66.
Leles da Costa Dias, F., Teixeira, A.L., Cerqueira Guimarães, H., Borges Santos, A.P., Rios Fonseca Ritter, S., Barbosa Machado, J.C., Tonidandel Barbosa, M., Caramelli, P., 2019. Prevalence of late-life depression and its correlates in a community-dwelling low-educated population aged 75+ years: The Pietà study. Journal of Affective Disorders 242, 173-179.
Leung, Y.-Y., Png, M.-E., Conaghan, P., Tennant, A., 2014. A Systematic Literature Review on the Application of Rasch Analysis in Musculoskeletal Disease — A Special Interest Group Report of OMERACT 11. The Journal of Rheumatology 41, 159-164.
Lundgren, Å.N., Tennant, A., 2011. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. Journal of rehabilitation medicine 43, 884-891.
Medvedev, O.N., Turner-Stokes, L., Ashford, S., Siegert, R.J., 2018. Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. Journal of rehabilitation medicine 50, 420-428.
Naushad, N., Dunn, L.B., Muñoz, R.F., Leykin, Y., 2018. Depression increases subjective stigma of chronic pain. Journal of Affective Disorders 229, 456-462.
Ng, T.P., 2018. Old age depression: worse clinical course, brighter treatment prospects? The Lancet Psychiatry.
Niu, L., Jia, C., Ma, Z., Wang, G., Yu, Z., Zhou, L., 2018. Validating the Geriatric Depression Scale with proxy-based data: A case-control psychological autopsy study in rural China. Journal of Affective Disorders 241, 533-538.
Perera, J.K., Rosenblat, J.D., Flint, A.J., 2017. Dementia with Lewy bodies presenting as psychotic depression. Australian & New Zealand Journal of Psychiatry 51, 1160-1161.
Pocklington, C., Gilbody, S., Manea, L., McMillan, D., 2016. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. International journal of geriatric psychiatry 31, 837-857.
Rash, G., 1960. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
Read, J.R., Sharpe, L., Modini, M., Dear, B.F., 2017. Multimorbidity and depression: A systematic review and meta-analysis. Journal of Affective Disorders 221, 36-46.
Reynolds, K., Pietrzak, R.H., El-Gabalawy, R., Mackenzie, C.S., Sareen, J., 2015. Prevalence of psychiatric disorders in US older adults: findings from a nationally representative survey. World Psychiatry 14, 74-81.
Sachdev, P.S., Brodaty, H., Reppermund, S., Kochan, N.A., Trollor, J.N., Draper, B., Slavin, M.J., Crawford, J., Kang, K., Broe, G.A., 2010. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. International Psychogeriatrics 22, 1248-1264.

Sheikh, J.I., Yesavage, J.A., 1986. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist: The Journal of Aging and Mental Health 5, 165-173.
Sjöberg, L., Karlsson, B., Atti, A.-R., Skoog, I., Fratiglioni, L., Wang, H.-X., 2017. Prevalence of depression: Comparisons of different depression definitions in population-based samples of older adults. Journal of Affective Disorders 221, 123-131.
Smith Jr, E.V., 2002. Understanding Rasch measurement: Detecting and evaluating the impact of multidimenstionality using item fit statistics and principal component analysis of residuals. Journal of applied measurement.
Svicher, A., Romanazzo, S., De Cesaris, F., Benemei, S., Geppetti, P., Cosci, F., 2019. Mental Pain Questionnaire: an Item Response Theory analysis. Journal of Affective Disorders 249, 226-233.
Tang, W.K., Wong, E., Chiu, H.F., Lum, C., Ungvari, G.S., 2005. The Geriatric Depression Scale should be shortened: results of Rasch analysis. International journal of geriatric psychiatry 20, 783-789.
Taouk, M., Meiser, B., Hadzi-Pavlovic, D., Dudley, M., Mitchell, P.B., 2017. Development of a potential screening measure for adolescent depression. The Australian and New Zealand journal of psychiatry 51, 624-633.
Tennant, A., Conaghan, P.G., 2007. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care & Research 57, 1358-1362.
Teo, A.R., Marsh, H.E., Forsberg, C.W., Nicolaidis, C., Chen, J.I., Newsom, J., Saha, S., Dobscha, S.K., 2018. Loneliness is closely associated with depression outcomes and suicidal ideation among military veterans in primary care. Journal of Affective Disorders 230, 42-49.
Tsoi, K.K.F., Chan, J.Y.C., Hirai, H.W., Wong, S.Y.S., 2017. Comparison of diagnostic performance of Two-Question Screen and 15 depression screening instruments for older adults: Systematic review and meta-analysis. British Journal of Psychiatry 210, 255-260.
Van Marwijk, H., Wallace, P., de Bock, G.H., Hermans, J., Kaptein, A.A., Mulder, J.D., 1995. Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. Br J Gen Pract 45, 195-199.
Wang, H.R., Cho, H., Kim, D.-J., 2018. Prevalence and correlates of comorbid depression in a nonclinical online sample with DSM-5 internet gaming disorder. Journal of Affective Disorders 226, 1-5.
Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., Leirer, V.O., 1982. Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research 17, 37-49.