

# Psychometric Validation of Geriatric Depression Scale – Short Form among Bengali-Speaking Elderly from a Rural Area of West Bengal: Application of Item Response Theory

Arista Lahiri<sup>1</sup>, Arup Chakraborty<sup>2</sup>

<sup>1</sup>Senior Resident, Department of Community Medicine, College of Medicine and Sagore Dutta Hospital, Kamarhati,

<sup>2</sup>Associate Professor, Department of Community Medicine, Medical College, Kolkata, West Bengal, India

## Abstract

**Background:** The geriatric depression scale – short form (GDS-SF) considered an important preliminary screening tool, requires translation in different regional languages and validation, to become of utmost use in Indian context especially in the field level. **Objective:** The current study aimed to evaluate the psychometric validity of the GDS-SF translated into Bengali among rural elderly population. **Methods:** The 15-item GDS-SF translated to Bengali language was administered to 206 Bengali speaking geriatrics selected maintaining predecided inclusion and exclusion criteria from villages under a rural block in West Bengal. Latent trait modelling was used to evaluate the psychometric properties of this translated tool. Differential item functioning (DIF) was assessed to measure invariance. **Results:** The mean age of the participants was 68.77 years (standard deviation 6.81 years). Majority being female (57.77%), Hindu (87.38%), from a joint family background (90.78%). The highest discrimination was observed with item 8 (coefficient 3.682,  $P < 0.001$ ) followed by item 14 (coefficient 3.020,  $P < 0.001$ ). Question 2 had least coefficient for difficulty ( $-1.344$ ,  $P = 0.013$ ) while item 15 had highest (0.775,  $P = 0.001$ ). The questionnaire provided maximum information (discrimination) around mean value of latent trait. The total cutoff score of 5 was related nearly to mean latent trait ( $-0.111$ ). Items 10 and 13 showed consistent DIF across different demographic groups. **Conclusion:** Psychometric properties of GDS-SF (Bengali) established overall construct and content validity of the tool in this community-based study. Despite some degree of DIF the tool can be used as a preliminary screening method in rural community.

**Key words:** Geriatric depression scale, geriatrics, item response theory, reliability, short form, validity

## INTRODUCTION

Worldwide the proportion of geriatric people is growing rapidly. As per the World Health Organization (WHO) elderly population worldwide will go up to 22% by 2050 from 12% estimated in 2015. They have also identified that mental illnesses are one of the very significant hindrances at achieving a disability-free life among the elderly (>60 years).<sup>[1-3]</sup> Estimates suggest around 7% of the elderly suffer unipolar depression.<sup>[4]</sup>

India is being second-most populated country in the world, also is home to its large geriatric population across country. Grover<sup>[5]</sup> reviewed different Indian literatures dealing with depression among elderly. While in the community-based studies the prevalence of geriatric depression ranges from ~8% to ~62%, the clinic-based studies reported ~42% to ~72%.<sup>[5]</sup> Geriatric depression is challenging in particular for primary

care setting, since it forms a greater proportion among different mental morbidities, yet are mostly under-diagnosed.<sup>[4,6]</sup> One way to combat these challenges is with field-level screening under primary-care structure. For this purpose, a proper screening tool or scale is of essence. The geriatric depression scale – short form (GDS-SF) has been an efficient tool for screening depression among elderly.<sup>[7-9]</sup>

**Address for correspondence:** Dr. Arup Chakraborty,  
240, Golpukur Road, Baruipur, 24 Parganas (South), Kolkata - 700 144,  
West Bengal, India.  
E-mail: [dr.arupchakraborty@gmail.com](mailto:dr.arupchakraborty@gmail.com)

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** [WKHLRPMedknow\\_reprints@wolterskluwer.com](mailto:WKHLRPMedknow_reprints@wolterskluwer.com)

**Submitted:** 07-Apr-2019

**Revised:** 20-Jul-2019

**Accepted:** 27-Apr-2020

**Published:** 16-Jun-2020

**How to cite this article:** Lahiri A, Chakraborty A. Psychometric validation of geriatric depression scale – Short form among Bengali-speaking elderly from a rural area of West Bengal: Application of item response theory. *Indian J Public Health* 2020;64:109-15.

### Access this article online

#### Quick Response Code:



**Website:**  
[www.ijph.in](http://www.ijph.in)

**DOI:**  
10.4103/ijph.IJPH\_162\_19

The GDS-SF has not been validated in Bengali yet. There has been many studies attempting to validate the scale in Hindi,<sup>[5]</sup> Tamil,<sup>[10]</sup> and Gujarati<sup>[11]</sup> and also in several other foreign languages such as Portuguese, Serbian, Iranian, Korean, and Nepali.<sup>[12-18]</sup> In several previous attempts at validating the GDS-SF in different languages the authors identified multifactor structure of the instrument.<sup>[13,15-18]</sup> Majority of them have addressed the issue with reliability estimates and discriminant validity through classical test theory assumptions. In assessing health-outcomes by self-report tools the psychometric properties pertaining to measurement of the latent trait holds the key about usefulness of the particular tool.<sup>[19]</sup> Among the researchers attempting to validate the tool, only few considered the psychometric property through the measurement of latent trait and construct. In this article, the authors attempted to evaluate the latent trait properties of the GDS-SF translated to Bengali and its usefulness for screening when administered among elderly in the rural areas.

## MATERIALS AND METHODS

### Study population and sampling

The current study was conducted among the Bengali-speaking geriatrics ( $\geq 60$  years) residing in eleven villages under the service area of randomly selected three subcentres under Barasat-II rural block of West Bengal. The number of subjects to be recruited from each village was calculated by probability proportionate to size (PPS) method. Segmental design was followed for recruitment of the participants from each village with the predecided criteria to select one elderly randomly from each selected household. Each village was divided into four quadrants and households were selected from each quadrant through PPS method. Participants having any diagnosed neurological disorder (e.g., seizure, Parkinsonism and tremors, paraplegia) were excluded. Bed-ridden and terminally ill geriatrics were also excluded from the study. The optimum sample size (taking prevalence of outcome 50%) calculated at 5% precision level with 80% power, a design effect of 2 and accounting for 10% nonresponse and partial-response was approximately 212. Since the objective was examining psychometric validity of the questionnaire, the minimum sample: item ratio was considered a minimum of 10:1 which yielded a minimum sample size of 150. After triangulating with the outcome prevalence-based calculation of sample size, to the item-based calculation, 212 was considered optimum sample size. After interviewing 212 participants during June–December, 2017 and then excluding the partial-responses obtained, ultimately, a total of 206 completed responses were considered for analysis.

### Study tool, translation and back-translation, and the survey

The study tool comprised of a translated version of the GDS-SF (also called GDS-15),<sup>[20]</sup> preceded by semi-structured questions on basic socio-demographic and clinical variables pertaining to the geriatrics. The GDS was developed as a self-report clinical screening tool for identifying depressive

symptoms among the elderly with a “yes/no” response pattern to the questions. The original scale comprised of 30 items in English.<sup>[21]</sup> For better feasibility a 15-item questionnaire was developed, which can be completed quickly, making the shorter version ideal for people with easy fatigue or limited ability to concentrate for longer periods of time.<sup>[8,22-24]</sup> Of the 15 items in the shorter version, i.e., GDS-SF, 10 indicate the presence of depression when answered positively while the remaining 5 (question numbers 1, 5, 7, 11, 13) are indicative of depression when answered negatively. For a response in favor of depression, 1 point is given and summing all the responses a total score is calculated. Score of up to 4 is considered normal, while a score of 12 or more almost always signify severe depression.<sup>[20,23]</sup>

The GDS-15 was translated to Bengali and then back-translated to English with the help of two different experts in the languages. Agreement among the original and the back-translated version was calculated for each of the items in the scale. Having accepted level of agreement set at 0.8 for individual items,<sup>[25]</sup> three items did not meet the criteria. These three items were then again put through the same translation and back-translation process with the help of different experts. On reaching the desired level of statistical agreement, the final Bengali version was accepted for the current study.

The participants were explained about the current study and the study tool before administering the questionnaire. The translated GDS, 15 questionnaires, was then administered with the help of trained interviewers (trained regarding the study tool in light of the objectives of the study) among the selected participants after obtaining consent. The Bengali-translated version of the questionnaire was used for the survey. Each of the participants was interviewed separately and their responses to the questions were noted.

### Psychometric validity

Item response theory (IRT) was used to measure the parameters for establishing validity in R 3.5.3 (The R Foundation, Vienna, Austria) with packages “lrm” and “TAM.”<sup>[26,27]</sup> Based on the method of modified parallel analysis, as put forward by Drasgow and Lissak to identify dimensionality of scale with dichotomous items using tetrachoric correlation matrix, “unidimTest” function (“lrm” package) established a single significant dimension for the data.<sup>[26,28]</sup> Therefore, uni-dimensional IRT models were considered appropriate. As seen with many of the health research-related self-report questionnaire, the responses are often not colluded with guessing contrary to those in other tests like an examination with multiple choice questions. Although it may be argued that, since items in a scale essentially belong to a single continuum of measurement, and their response choices are also similar (dichotomous), therefore they can be considered having similar discrimination throughout, with the exception of difficulty. At this point, two approaches emerge for analysis: 1-parameter logistic model (1-PL) and 2-parameter logistic model (2-PL), with

the only differentiating factor – “discrimination” being considered equal along all the items in the 1-PL model. It can be further conceptualized that, in health-related questionnaires, discrimination parameter is not likely to be equal along the scale. The factors to be considered are the number of items and also the presence of reverse-coded items in a scale. Importantly for self-report questionnaires in health outcomes individual items tend to carry uneven discriminative ability. It is often because of this uneven discrimination the cutoffs in health-outcomes tend to vary. The 2-PL model appeared statistically appropriate compared to the 1-PL model as per maximum marginal likelihood estimation, Akaike Information Criterion, and Bayesian Information Criterion.<sup>[29]</sup> Considering all the factors 2-PL model was selected. The items in the questionnaire were analyzed with the help of discrimination and difficulty coefficients, item, and test information functions. Infit and outfit statistics were calculated based on mean square statistic. A value of 0.5-1.5 was considered acceptable for measurement of the parameters in the 2-PL model.<sup>[30]</sup>

The test information function provides cumulative information, thus providing local reliability over different levels of latent

trait. Similarly, the test characteristics function provides different levels of discrimination over different level of the latent trait.<sup>[19,31,32]</sup> For graphical representations of test information and test characteristic functions [Figure 1], the latent trait, i.e., ability denoted by theta ( $\theta$ ), which is actually the continuum of depression measured through the survey, has been presented in the X-axis. In order to measure whether the items behave consistently or reliably across the different socio-demographic groups gender, religion, family type, marital status, and educational status were selected for analysis of invariance through differential item functioning (DIF) using logistic model. In the paradigm of latent trait modelling invariance helps in factoring out confounding and effect modification contributed by these variables.

### Ethical issues

The study was approved by the Institutional Ethics Committee, Medical College and Hospital, Kolkata (Ref. No.MC/Kol/IEC/Non-Spon/569/05-2017). Permission was taken from the Barasat II block administrative officials. The data collection was conducted only after obtaining written consent from the respondents. The survey was conducted maintaining anonymity and confidentiality of the participants. For, those who were noted to have a score indicative of depression were advised to visit a health-care facility with psychiatry clinic. However, no incentives whatsoever were offered to the respondents.

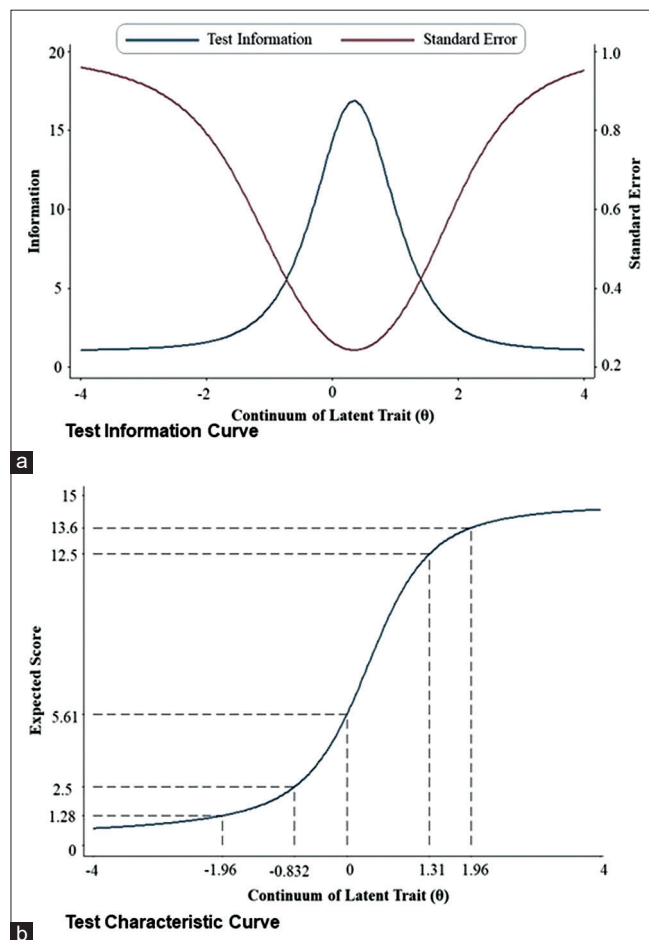
## RESULTS

### Background characteristics

The socio-demographic profile of the study participants is summarized in Table 1. The mean age of the participants was observed to be 68.77 years ( $\pm 6.81$  years). The youngest participant was aged 60 years and the oldest was 88 years old. Among the participants majority were female (57.77%), Hindu (87.38%), married (68.93%). Around 90.78% of the respondents were from a joint family background. Same proportion of the respondents was observed to be living with children. Majority had sedentary level of physical activity (50.49%). Among the respondents, 35.44% were illiterate, 29.61% had preprimary and 23.30% primary level of education. Among the elderly interviewed, 43.22% were addicted to tobacco chewing, 31.25% to smoked tobacco, and 13.25% to alcohol. The remaining 56.80% did not report to have any addictions.

### Two-parameter logistic item response theory model

The 2-PL model for 15-item GDS is shown in Table 2 along with the infit and outfit statistics. While the discrimination parameter differentiates the respondents on the items of the scale, the difficulty parameter does the opposite. Items 8 and 14 were observed to have high discrimination parameter, 3.682 (95% confidence interval [CI]: 2.285–5.079) and 3.020 (95% CI: 1.931–4.109) respectively and both were statistically significant. However, the difficulty parameter was higher for item 14 compared to item 8,



**Figure 1:** Test performance of geriatric depression scale – short form as per 2-parameter logistic model item response theory model. (a) Test Information Curve (b) Test Characteristic Curve

**Table 1: Socio-demographic characteristics of the study participants (n=206)**

Socio-demographic characteristics	n (%)
Sex	
Male	87 (42.23)
Female	119 (57.77)
Religion	
Hinduism	180 (87.38)
Islam	26 (12.62)
Marital status	
Married	142 (68.93)
Widowed	64 (31.07)
Type of family	
Nuclear	19 (9.22)
Joint	187 (90.78)
Physical activity	
Sedentary	104 (50.49)
Moderate	76 (36.89)
Heavy	26 (12.62)
Educational status	
Illiterate	73 (35.44)
Preprimary	61 (29.61)
Primary	48 (23.30)
Middle	23 (11.17)
Secondary	1 (0.49)
Living arrangement	
Living alone	19 (9.22)
Living with children	187 (90.78)
Addiction	
No addiction	117 (56.80)
Addicted to tobacco and/or alcohol	89 (43.20)

0.598 (95% CI: 0.376–0.805) and 0.482 (95% CI: 0.081–0.459), respectively. Item 1 documented a discrimination estimate of 2.698 (95% CI: 1.775–3.622). Discrimination estimates were comparable with that of item 1, for item 4 (2.640, 95% CI: 1.715–3.566), item 7 (2.634, 95% CI: 1.743–3.526). Although comparatively lower, but items 5 and 12 also reported a good discrimination with estimates being 2.397 (95% CI: 1.589–3.205) and 2.322 (95% CI: 1.500–3.143), respectively. The highest difficulty was observed with items 15 and 12 with estimates of 0.775 (95% CI: 0.310–1.241) and 0.764 (95% CI: 0.515–1.014), respectively.

Among reverse coded items, item 1 followed by items 7 and 5 had difficulty estimates of 0.522 (95% CI: 0.304–0.740), 0.329 (95% CI: 0.120–0.538), and 0.291 (95% CI: 0.076–0.506) in respective order. Two more items, i.e., item 11 and 13 were also reverse coded. The difficulty of these two items was 0.070 and 0.227, respectively. However, these estimates were not statistically significant. For item 9, equal response was noted in terms of yes and no. The discrimination and difficulty both for this item was low and the estimates were statistically insignificant. While overall the items proved to be having good discrimination, the difficulty was low with few items only showing higher level of difficulty.

**Table 2: Two-parameter logistic model of geriatric depression scale-short form**

Items	Response: Yes, n (%)		Discrimination		Difficulty		Outfit statistics**		Infit statistics**	
	Estimate	P	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate	P	Estimate	P
1. Are you basically satisfied with your life?*	138 (66.99)	<0.001	2.698 (1.775-3.622)	<0.001	0.522 (0.304-0.740)	<0.001	1.424	<0.001	0.939	<0.001
2. Have you dropped many of your activities and interests?	134 (65.05)	0.006	0.488 (0.139-0.837)	0.006	-1.344 (-2.406--0.282)	0.013	0.979	0.013	1.008	0.013
3. Do you feel that your life is empty?	99 (48.06)	<0.001	2.019 (1.337-2.702)	<0.001	0.047 (-0.175-0.269)	0.677	1.080	0.677	0.990	0.677
4. Do you often get bored?	104 (50.49)	<0.001	2.640 (1.715-3.566)	<0.001	-0.039 (-0.240-0.162)	0.702	1.186	0.702	0.983	0.702
5. Are you in good spirits most of the time?*	123 (59.71)	<0.001	2.397 (1.589-3.205)	<0.001	0.291 (0.076-0.506)	0.008	1.040	0.008	0.994	0.008
6. Are you afraid that something bad is going to happen to you?	76 (36.89)	<0.001	1.598 (1.048-2.148)	<0.001	0.482 (0.213-0.752)	<0.001	1.018	<0.001	0.999	<0.001
7. Do you feel happy most of the time?*	126 (61.17)	<0.001	2.634 (1.743-3.526)	<0.001	0.329 (0.120-0.538)	0.002	0.899	0.002	1.019	0.002
8. Do you often feel helpless?	82 (39.81)	<0.001	3.682 (2.285-5.079)	<0.001	0.270 (0.081-0.459)	0.005	0.926	0.005	1.018	0.005
9. Do you prefer to stay at home, rather than going out and doing new things?	103 (50.00)	0.801	0.039 (-0.262-0.340)	0.801	0.0001 (-7.049-7.049)	1.000	1.000	1.000	1.000	1.000
10. Do you feel you have more problems with memory than most?	85 (41.26)	<0.001	1.378 (0.885-1.871)	<0.001	0.341 (0.060-0.622)	0.018	1.071	0.018	0.991	0.018
11. Do you think it is wonderful to be alive now?*	108 (52.43)	<0.001	1.785 (1.179-2.390)	<0.001	0.070 (-0.165-0.305)	0.558	1.042	0.558	0.998	0.558
12. Do you feel pretty worthless the way you are now?	56 (27.18)	<0.001	2.322 (1.500-3.143)	<0.001	0.764 (0.515-1.014)	<0.001	0.861	<0.001	1.030	<0.001
13. Do you feel full of energy?*	94 (45.63)	<0.001	0.943 (0.540-1.347)	<0.001	-0.227 (-0.577-0.123)	0.203	1.013	0.203	1.001	0.203
14. Do you feel that your situation is hopeless?	63 (30.58)	<0.001	3.020 (1.931-4.109)	<0.001	0.591 (0.376-0.805)	<0.001	1.328	<0.001	0.976	<0.001
15. Do you think that most people are better off than you are?	74 (35.92)	<0.001	0.862 (0.473-1.250)	<0.001	0.775 (0.310-1.241)	0.001	0.984	0.001	1.008	0.001

\*Reverse scored items: Score for "Yes" response is 0 and "No" is 1, \*\*A value of 0.5-1.5 is considered acceptable. CI: Confidence interval

### Differential item functioning

The variability of these 15 items on the latent trait continuum with respect to different socio-demographic groups is shown in Table 3 with DIF. The present article deals with variability of response across different groups but not about the measurement of this variability. Table 3 shows the logistic test for DIF, which identifies statistically significant variability in response for each of the items. No items reported DIF as per different sex of the participants. However, item 1 showed statistically significant difference in response for different religions. This item also showed differential functioning on latent trait continuum (nonuniform) with respect to different marital status but was ultimately not significant statistically at constant trait level (uniform). Item 2 recorded only nonuniform DIF for educational status (illiterate vs. literate). Nonuniform DIF was also statistically significant with item 3 (family type and educational status), item 4 (family type), item 6 (marital status and family type), item 7 (family type), item 8 (educational status), item 12 (marital status), and item 13 (marital status and religion). Interestingly, for item 9, statistically significant differential functioning both nonuniform and uniform was noted with educational status of the participants. For item 10, the differential response was noted for difference in educational status only over some range of latent trait (nonuniform), but with different religion among the respondent the differential response was significant consistently. For item 11, different religions of the respondents incurred significant nonuniform and uniform DIF. Items 5, 14, and 15 did not show variability over these socio-demographic groups.

### Overall property of geriatric depression scale – short form

Test information curve (TIC) is shown in Figure 1a. The TIC plots the test information function, while the standard error curve is a pictorial measure of change in precision. It is evident from the TIC that the maximum information is obtained around the mean level of latent trait. Interestingly, the peak of information is not at the mean, rather is situated at a positive distance from the mean latent trait level. The information from the GDS-SF reaches minimum along both the tails of the curve that means as the latent trait level goes further away from the mean in either direction. The standard error of measurement is minimum at the peak information level and it rises as the information decreases in either side. Figure 1b depicts the test characteristic curve (TCC), which is basically the summation of all the item characteristics. This curve shows that discrimination is best around the mean (-0.832–1.96). The mean level of latent trait corresponds with expected score of 5.61, i.e., more than 5 (the cutoff). The test below the trait level of -0.832 and above 1.96 reaches a plateau which signifies that in these ranges the overall discrimination gets poorer. As per the cutoff, 57.8% of the participants were found to be depressed.

### DISCUSSION

The traditional methods used for validity analysis for example factor analysis, takes into account the presence of latent traits

**Table 3: Differential item functioning of geriatric depression scale-short form items on the basis of logistic test**

Items	$\chi^2 (P)$														
	Sex			Religion			Marital status			Family type			Educational status		
	Nonuniform	Uniform		Nonuniform	Uniform		Nonuniform	Uniform		Nonuniform	Uniform		Nonuniform	Uniform	
1	1.350 (0.245)	0.980 (0.322)		5.680 (0.017)	5.800 (0.016)		6.560 (0.010)	1.800 (0.180)		2.820 (0.093)	4.670 (0.031)		<0.001 (0.980)	4.790 (0.029)	
2	3.640 (0.056)	2.770 (0.096)		0.010 (0.927)	15.680 (<0.001)		0.090 (0.771)	1.610 (0.205)		2.560 (0.110)	0.980 (0.321)		4.600 (0.032)	0.310 (0.575)	
3	2.290 (0.130)	14.910 (<0.001)		0.300 (0.583)	10.170 (0.001)		0.280 (0.598)	0.010 (0.919)		4.120 (0.042)	2.330 (0.127)		5.550 (0.019)	1.510 (0.220)	
4	0.040 (0.839)	0.010 (0.903)		3.300 (0.069)	0.920 (0.338)		0.020 (0.897)	10.590 (0.001)		8.850 (0.003)	0.300 (0.585)		-- <sup>a</sup>	-- <sup>a</sup>	
5	0.110 (0.737)	4.600 (0.032)		0.530 (0.467)	5.530 (0.019)		0.020 (0.883)	0.160 (0.686)		0.280 (0.596)	1.490 (0.223)		0.750 (0.387)	1.060 (0.303)	
6	1.900 (0.168)	0.320 (0.573)		0.060 (0.801)	14.100 (<0.001)		11.280 (0.001)	0.680 (0.409)		4.760 (0.029)	0.940 (0.332)		3.600 (0.058)	1.260 (0.262)	
7	1.830 (0.176)	0.010 (0.908)		0.670 (0.414)	0.820 (0.364)		0.100 (0.751)	0.790 (0.375)		5.140 (0.023)	<0.001 (0.976)		0.580 (0.446)	0.080 (0.784)	
8	0.610 (0.433)	1.480 (0.225)		0.040 (0.846)	13.840 (<0.001)		0.050 (0.831)	<0.001 (0.979)		3.200 (0.074)	0.010 (0.910)		5.090 (0.024)	2.410 (0.120)	
9	0.720 (0.396)	0.060 (0.810)		3.110 (0.078)	1.090 (0.296)		0.020 (0.883)	6.010 (0.014)		0.270 (0.607)	0.340 (0.560)		5.790 (0.016)	6.470 (0.011)	
10	0.610 (0.435)	0.940 (0.332)		9.910 (0.002)	4.230 (0.040)		1.220 (0.270)	2.680 (0.101)		<0.001 (0.965)	1.570 (0.210)		9.760 (0.002)	0.420 (0.515)	
11	1.970 (0.160)	9.300 (0.002)		10.380 (0.001)	5.170 (0.023)		0.010 (0.910)	0.030 (0.864)		3.800 (0.051)	0.900 (0.343)		2.570 (0.109)	5.490 (0.019)	
12	0.490 (0.486)	4.500 (0.034)		0.400 (0.530)	10.450 (0.001)		6.240 (0.013)	0.870 (0.351)		-- <sup>a</sup>	-- <sup>a</sup>		0.730 (0.393)	5.850 (0.016)	
13	1.010 (0.314)	20.630 (<0.001)		15.780 (<0.001)	2.610 (0.106)		5.850 (0.016)	1.290 (0.257)		2.390 (0.122)	4.140 (0.042)		2.610 (0.106)	14.880 (<0.001)	
14	0.590 (0.441)	3.040 (0.081)		0.180 (0.673)	0.830 (0.361)		0.140 (0.712)	0.550 (0.459)		-- <sup>a</sup>	-- <sup>a</sup>		0.210 (0.651)	0.530 (0.465)	
15	1.360 (0.244)	3.720 (0.054)		2.380 (0.123)	0.440 (0.507)		0.040 (0.844)	0.090 (0.768)		0.460 (0.496)	1.320 (0.251)		1.360 (0.243)	2.710 (0.100)	

<sup>a</sup>Value could not be computed

in a questionnaire, but considers the responses observed on a continuous scale which is not always appropriate, as in the case of GDS-SF. Another commonly employed approaches to handle categorical, typically nominal responses in validation studies are correspondence and cluster analyses techniques. However, the basic problem with these is that they can group the responses, but are not able to measure the underlying trait that is actually measured. Hays *et al.*,<sup>[19]</sup> thus emphasized on using IRT for evaluating validity in outcome measure scales, since it can actually measure along the continuum of latent trait in the questionnaire. From the construct of the questionnaire (yes/no response), it was evident that a binary item-response model would be appropriate for the analysis. Latent trait modelling in contrast to classical test theory provides the computational means to address the variability of reliability and discrimination along the changing continuum of trait level for validity of a scale.

The current article did not seek factors but identified a single dimension underlying the latent trait continuum of geriatric depression being measured by GDS-15. The results obtained in the current study reveals that the scale performs best at just above mean level of latent trait. The local reliability<sup>[31]</sup> is best at a positive distance from mean. However, under the assumptions of classical test theory the GDS-15 was found to be highly reliable as well (Cronbach's alpha: 0.87). Brown and Schinka<sup>[7]</sup> reported a high internal consistency and reliability. The Portuguese,<sup>[16]</sup> Iranian,<sup>[13]</sup> and Serbian<sup>[18]</sup> version reported high values of Cronbach's alpha (>0.9), while the Gujarati version<sup>[11]</sup> had an alpha value of >0.8, which is also a very acceptable value. On the other hand, the discrimination as analysed in the current article is acceptable within the range of  $\pm 1.96$  around mean in the latent trait continuum, though within the range  $-0.832$  and  $1.31$  the discrimination is optimum and hence the steep stem in the TCC. Comparable to the results in the current article, the Chinese version also reported satisfactory discrimination.<sup>[33]</sup>

When the TIC and TCC are compared, it is evident that with loss of local reliability outside the mentioned range ( $\pm 1.96$  around mean) the discrimination obtained from the test gets poorer. TCC alone however can provide information regarding the best cutoff for optimum discrimination. In the validation of the Iranian version, the authors used receiver operating characteristic analysis to identify the cutoff and reported a value of 7/8 to have optimum specificity and sensitivity.<sup>[13]</sup> Similar cutoff was suggested by the authors validating the Tamil version of the scale.<sup>[10]</sup> The current study reveals the cutoff to be 5.61 at the mean trait level. However, 5.61 is not a practical score from the tool, thus either the cutoff should be at 6 or at 5. For GDS-SF the cutoff is set at a total score of 5, which appears to be slightly below the mean trait level in the case of the Bengali version. However, this is acceptable as the objective of using this tool in the field setting is for effective screening.

The positive direction in the trait continuum implies a stronger probability of having depression. After the score of

13.6 i.e., 1.96 in the trait continuum the TCC loses its slope, thereby defining the loss of discrimination. In other words, it can be conceptualized that after this level there is no further discrimination in terms of having depression or those after this level are almost certain to have depression. However, at the trait level of 1.31 (expected score 12.5) the TCC loses its steep slope and the rate of change in slope that is discrimination ability starts to decrease. Therefore, in a broader perspective 12.5 rather than 13.6 can be treated as the cutoff to be almost certain to have depression. Again 12.5 is not a practical score in the scale, so the authors take 13. However, the scale guidelines<sup>[8,22,34]</sup> suggest that a score of 12 or more is almost always suggestive of depression. However, from a public health point of view, the use of this instrument at the field level will warrant a single cut-off of 5 for screening utility.

Finally, the issue of item invariance across groups needs to be considered, as it implies a cardinal pillar in the structural validity of the tool. Almost all the 15 items in the instrument were invariant across different socio-demographic groups, implying that the questions did not carry different meanings across different groups of elderly. However, some degree of DIF was noted over varying trait level. Midden and Mast<sup>[35]</sup> examined the DIF with respect to cognitive impairment status of the patients and found out that there was satisfactory level of item invariance across groups.

The current study, emphasizes on the fact that the Bengali translated version of the GDS-SF can be used as a field screening instrument among the rural elderly for the identification of probable depression. However, the article does not provide information regarding the validity among those with neurological abnormalities or otherwise diagnosed neuropsychiatric disorders. Therefore, if used at field level, it should be interpreted cautiously. Some items could not be tested for invariance for some socio-demographic variables at the current sample size. A higher power could have yielded measurements in this regard. Nevertheless, a community-based validity analysis adds to the strengths of the current study alongside the use of more rational analytical method, i.e., IRT for establishing the validity of the Bengali translation of the GDS-SF.

## CONCLUSION

It is recommended by the authors to use the version as a field-level screening tool. The scale-dependent outcome of probable depression, can definitely be regarded as the hidden portion of the morbidity – the subclinical depression, which would also comprise of clinical or overt depression manifestations, diagnosed at specialist facilities following referral after field-screening. A cutoff value of 5 can be used for screening purposes. In a nut-shell, the Bengali version of GDS-SF did show reliable and discriminative psychometric properties. Items on DIF analysis showed acceptable level of consistency (or invariance) across socio-demographic groups among geriatrics.

## Acknowledgment

The authors would like to acknowledge the study participants, without their co-operation the study would not have been possible. The authors would also like to acknowledge the ASHAs and ANMs of the study villages under Barasat-II block. Their active support made the survey component get completed within due time.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

- World Health Organization. Proposed Working Definition of an Older Person in Africa for the MDS Project. Geneva: World Health Organization. Available from: <http://www.who.int/healthinfo/surveys/ageingdefolder/en>. [Last accessed on 2018 Dec 23].
- World Health Organization. Ageing and Life Course. Geneva: World Health Organization. Available from: <http://www.who.int/ageing/en>. [Last accessed on 2018 Dec 23].
- Mental Health of Older Adults. Available from: <https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults>. [Last accessed on 2018 Dec 23]
- World Health Organization. Mental Health and Older Adults. Geneva: World Health Organization. Available from: <http://www.who.int/mediacentre/factsheets/fs381/en>. [Last accessed on 2017 Oct 15].
- Grover S, Malhotra N. Depression in elderly: A review of Indian research. *J Geriatr Ment Health* [serial online] 2015;2:4-15. Available from: <http://www.jgmh.org/text.asp?2015/2/1/4/161376>. [Last accessed on 2019 Feb 13].
- Pilania M, Bairwa M, Kumar N, Khanna P, Kurana H. Elderly depression in India: An emerging public health challenge. *Australas J Geriatr* 2013;6:107-11.
- Brown LM, Schinka JA. Development and initial validation of a 15-item informant version of the geriatric depression scale. *Int J Geriatr Psychiatry* 2005;20:911-8.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, *et al*. Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatr Res* 1982-1983;17:37-49.
- Burke WJ, Roccaforte WH, Wengel SP. The short form of the geriatric depression scale: A comparison with the 30-item form. *Top Geriatr* 1991;4:173-8.
- Sarkar S, Kattimani S, Roy G, Premarajan KC, Sarkar S. Validation of the Tamil version of short form geriatric depression scale-15. *J Neurosci Rural Pract* 2015;6:442-6.
- Desai N, Shah S, Sharma E, Mishra A, Mehta K. Validation of Gujarati version of 15-item geriatric depression scale in elderly medical outpatients of general hospital in Gujarat. *Int J Med Sci Public Health* 2014;3:1453.
- Bae JN, Cho MJ. Development of the Korean version of the geriatric depression scale and its short form among elderly psychiatric patients. *J Psychosom Res* 2004;57:297-305.
- Malakouti SK, Fatollahi P, Mirabzadeh A, Salavati M, Zandi T. Reliability, validity and factor structure of the GDS-15 in Iranian elderly. *Int J Geriatr Psychiatry* 2006;21:588-93.
- Validation Malay Geriatric Depression Scale. Available from: <http://www.priory.com/psych/MalayGDS.htm>. [Last access on 2017 Dec 10].
- Gautam R, Houde SC. Geriatric depression scale for community-dwelling older adults in Nepal. *Asian J Gerontol Geriatr* 2011;6:93-9.
- Apóstolo J, Loureiro L, Reis I, Silva I, Cardoso D, Sfetcu R. Contribution to the adaptation of the geriatric depression scale 15 into Portuguese. *Rev Enferm Referência* 2014;4:65-73.
- Ştefan AM, Băban A. The Romanian version of the geriatric depression scale: Reliability and validity. *Cogn Creier Comportament* 2017;21:175.
- Stolić D, Jović J, Bukumirić Z, Rančić N, Stolić M, Ignjatović-Ristić D. The Serbian version of the geriatric depression scale: Reliability, validity and psychometric features among the depressed and non-depressed elderly. *Engrami* 2015;37:51-64.
- Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21<sup>st</sup> century. *Med Care* 2000;38 Suppl 9:II28-42.
- Geriatric Depression Scale Short Form English Scoring. Available from: <https://web.stanford.edu/~yesavage/GDS.english.short.score.html>. [Last accessed on 2018 Dec 25]
- Geriatric Depression Scale Long Form English. Available from: <https://web.stanford.edu/~yesavage/GDS.english.long.html>. [Last accessed on 2018 Dec 25].
- Geriatric Depression Scale Short Form (GDS-SF). Psych Congress. Available from: <https://www.psychcongress.com/saunders-corner/scales-screeners/depression/geriatric-depression-scale-short-form-gds-sf>. [Last accessed on 2018 Oct 08].
- Sheikh JI, Yesavage JA. Geriatric depression scale (GDS): Recent evidence and development of a shorter version. *Clin Gerontol* 1986;5:165-73.
- Geriatric Depression Scale. Available from: <http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/geriatric-depression.aspx>. [Last accessed on 2018 Dec 25].
- McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med* 2012;22:276-82.
- CRAN Package LTM. Available from: <https://cran.r-project.org/web/packages/ltm/index.html>. [Last accessed on 2020 Mar 20].
- CRAN Package TAM. Available from: <https://cran.r-project.org/web/packages/TAM/index.html>. [Last accessed on 2020 Mar 20].
- Drasgow F, Lissak RI. Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *J Appl Psychol* 1983;68:363-73.
- Test Analysis Modules: Marginal Maximum Likelihood Estimation in TAM: Test Analysis Modules. Available from: <https://rdrr.io/cran/TAM/man/tam.mml.html>. [Last accessed on 2020 Mar 20].
- What do Infit and Outfit, Mean-square and Standardized Mean? Available from: <https://www.rasch.org/rmt/rmt162f.htm>. [Last accessed on 2020 Mar 20].
- Fan X. Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educ Psychol Meas* 1998;58:357-81.
- Chiesi F, Primi C, Pigliautile M, Ercolani S, Staffa MC della, Longo A, *et al*. The local reliability of the 15-item version of the geriatric depression scale: An item response theory (IRT) study. *J Psychosom Res* 2017;96:84-8.
- Chiu HF, Lee HC, Wing YK, Kwong PK, Leung CM, Chung DW. Reliability, validity and structure of the Chinese geriatric depression scale in a Hong Kong context: A preliminary report. *Singapore Med J* 1994;35:477-80.
- Geriatric Depression Scale. Available from: <https://web.stanford.edu/~yesavage/GDS.html> [Last accessed on 2018 Oct 08].
- Midden AJ, Mast BT. Differential item functioning analysis of items on the Geriatric Depression Scale-15 based on the presence or absence of cognitive impairment. *Aging Ment Health* 2017;22:1136-42.