Journal of Affective Disorders 227 (2018) 471–476

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Does the 15-item Geriatric Depression Scale function differently in old people with different levels of cognitive functioning?

Francesca Chiesi[a,*], Caterina Primi[a], Martina Pigliautile[b], Marta Baroni[b], Sara Ercolani[b], Lucia Paolacci[b], Virginia Boccardi[b], Patrizia Mecocci[b]

[a] Department of Neuroscience, Psychology, Drug, and Child's Health (NEUROFARBA), Section of Psychology, University of Florence, Italy
[b] Institute of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Italy

ABSTRACT

Background: The 15-item version of the Geriatric Depression Scale (GDS-15) is widely employed to screen depression among elderly but little is known about the scale functioning in cognitively impaired individuals when compared to normal ones. The aim of the current study is to investigate Differential Item Functioning (DIF) across groups of older people that differ in terms of cognitive functioning applying Item Response Theory (IRT)-based analyses.

Methods: Data from an Italian multi-centric clinical-based study on cognitive impairment and dementia in old people were employed (N = 1903; Age: M = 77.33, SD = 7.05, 62% women). All the participants underwent a comprehensive evaluation (including clinical examination, laboratory screening, neuroimaging, and cognitive and behavioral assessments) and they were assigned to three different groups on the basis of their cognitive functioning (normal, mild cognitive impairment, cognitive impairment)

Results: Two items showed uniform DIF but their differential functioning does not propagate to the GDS-15 total scores in such a way that a differential interpretation is needed

Limitations: Whereas an advantage of the study is the large sample size, the relatively small size of the mild cognitive impairment group might reduce the stability of the present results

Conclusions: Since a screening tool for elderly is intended to apply to everyone in the target population, the current findings support the clinical utility of the GDS-15 as screening tool for depression.

1. Introduction

The 15-item Geriatric Depression Scale (GDS-15; Yesavage and Sheikh, 1986) is a widely employed screening test that offers an added value in the primary care detection of late-life depression. Indeed, it requires short time and effort to administer and it has good psychometric properties (for reviews, see Azulai and Walsh, 2015; Mitchell et al., 2010; Pocklington et al., 2016).

Regardless the extensive use (for a recent review, see Pocklington et al., 2016), it exists a debate on the possibility to administer the scale to people with a cognitive impairment, which is a very common and comorbid to depression disease in old age (for a review, see Wang and Blazer, 2015). On one hand, Wancata et al. (2006) affirmed that the scale should not be used with persons with marked cognitive impairment and Burke and colleagues (Burke et al., 1989) stated that only people with levels of 0 and 1 (i.e., cognitively intact and mildly demented individuals, respectively) on the Clinical Dementia Rating (CDR; Berg, 1984) were able to complete the test. On the other hand, Lach et al. (2010) built evidence for the use of the GDS-15 in populations that include people with mild to moderate dementia. Additionally, several studies (Conradsson et al., 2013; Jongenelis et al., 2005; McGivney et al., 1994; Smalbrugge et al., 2008) attested that the GDS-15 seems to be valid for people with mild cognitive impairments —assessed by the Mini-Mental State Examination (MMSE; Folstein et al., 1975)—, but it is unclear if the scale can be used with people with lower cognitive functioning (Conradsson et al., 2013).

This brief review suggests that further research is needed to understand whether cognitive impairments might produce biased responses or prevent the correct use of the GDS-15 (Luppa et al., 2012; Watson and Pignone, 2003). Specifically, it is important to ensure that other variables different from depression (i.e., the construct that the test seeks to measure) do not have an impact on the total test score. Since cognitive impairment might be a respondent's characteristics that produces biases on the GDS-15 total score, it is important to establish

* Correspondence to: Department of Neuroscience, Psychology, Drug, and Child's Health (NEUROFARBA), Section of Psychology, via di San Salvi 12- Padiglione 26, 50135 Firenze, Italy.
E-mail address: francesca.chiesi@unifi.it (F. Chiesi).

https://doi.org/10.1016/j.jad.2017.11.045
Received 28 January 2017; Received in revised form 9 September 2017; Accepted 11 November 2017
Available online 13 November 2017
0165-0327/

empirically whether test items work in the same way across people with different cognitive functioning.

From a psychometric point of view, this issue is adequately addressed by Item Response Theory (IRT) that allows the assessment of Differential Item Functioning (DIF; Embretson and Reise, 2000; Reise and Waller, 2009). DIF analysis is used to study the performance of items in scales, and it examines whether or not the likelihood of an item endorsement is equal across subgroups (e.g., people with different levels of cognitive functioning), which are matched on the trait measured (e.g., depression). For example, a randomly selected individual with a normal cognitive functioning and a specific level of depression and a randomly selected individual with an impaired cognitive functioning but the same level of depression should have the same likelihood of endorsing an item measuring depression. This aspect is of particular relevance to screening tools because, as Zumbo (2003) pointed out, the presence of DIF might produce a systematic bias in the total test scores and, as a consequence, interpretations based on cut-off scores might be biased (Hidalgo, Galindo-Garre, and Gómez-Benito, 2015; Jones and Raju, 2000; Stark et al., 2004).

Starting from this premise, the current study aimed at further investigating potential DIF in the GDS-15 applying IRT. To the best of our knowledge, only one study addressed this issue investigating differential functioning of the GDS-15 items applying Rash analyses (Tang et al., 2005). Comparing three different cognitive functioning groups (defined using the MMSE), Tang et al. (2005) investigated DIF looking at item location parameters, i.e., if one group is consistently more likely than another to endorse an item, and they reported evidence of no DIF. In the current study, to ascertain if there are biases in the measurement process among individuals with different levels of cognitive function, we studied also item discrimination parameters that can be viewed as a significant group by trait interaction, i.e., if one group is more likely to endorse an item at certain levels of the trait, while the other group is more likely to endorse the item at other levels. Additionally, we focused on the assignment of individual to each studied group. As stated above, it is difficult to derive clear indications from the literature about the use of the GSD-15 in case of cognitive impairment because cognitive impairment was measured referring mainly to a single test (e.g., the CDR or the MMSE) and a variety of inclusion criteria for cognitive functioning levels and subgroup definition have been used. Thus, in order to ensure a valid classification for the variable of interest (i.e., cognitive functioning) and go beyond the limitations of previous studies, all the participants of the current study were classified after a comprehensive evaluation, including clinical examination, laboratory screening, neuroimaging, and cognitive and behavioral assessments.

2. Methods

2.1. Participants and measure

Data were gathered from those collected at the Institute of Gerontology and Geriatrics University of Perugia according to the methodology of the ReGAl project (Rete Geriatrica Alzheimer—Geriatric Network on Alzheimer's disease), a large longitudinal Italian multicentre clinical-based study, promoted by the Italian Society of Gerontology and Geriatrics (SIGG), and focused on cognitive impairment and dementia in old people, as described elsewhere (Boccardi et al., 2016; Mariani et al., 2008). All experimental procedures were conducted in accordance with the guidelines in the Declaration of Helsinki and approved by the Ethics Committee of the Hospital-University of Perugia.

The study enrolled 1903 people (Age: $M = 77.33$, $SD = 7.05$, range 45–96; 62% women) from September 2011 to June 2014. The neuropsychological battery included the MMSE, as test of general cognition, and specific tests evaluating episodic memory (Babcock Story Recall test and the immediate and delayed recall of the Rey's Auditory Verbal Learning Test), language (Token test for verbal comprehension and the Category Fluency test for language production according to semantic cues), attention and executive functions (Visual Search test and the Letter Fluency test) and praxis (Copy Drawing test). For each test, details on administration procedures and Italian normative data for score adjustment for age and education as well as normality cut-off scores (95% of the lower tolerance limit of the normal population distribution) are available (Carlesimo et al., 1996; Spinnler and Tognoni, 1987). To avoid the underestimation of the level of functional capacity, informant based rating of functional status were carried out (Tabert et al., 2002) using the BADL (Katz et al., 1963) and the IADL scales (Lawton and Brody, 1969). In most of the cases, informants were spouses or relatives, who lived in the same household. BADL includes six activities: bathing, dressing, toileting, transferring, continence, and feeding. IADL includes eight activities: using telephone, shopping, meal preparation, housekeeping, laundry, use of transportation, self-administration of drugs, and handling finances. Because IADL items are often gender-specific, we considered not only the current ability to perform each item but also the potential capability in case of need. For BADL score ranges from 0 (total independence) to 6 (total dependence), and for IADL from 0 (total independence) to 8 (total dependence). The CDR was used to score dementia severity. Finally, the battery included the GDS-15.

General exclusion criteria were the presence of clinically severe psychiatric or systemic disease, severe sensory impairment (blindness, deafness), neurological conditions associated with severe cognitive impairment (i.e., severe dementia or advanced stages of Alzheimer disease), a history of alcohol or substance abuse or dependence, and head injury with loss of consciousness. Therefore, 598 cases were excluded from the initial pool of data.

The remaining 1305 cases were classified as no cognitive impairment (NCI), mild cognitive impairment (MCI), and cognitive impairment (CI) through the multidimensional assessment derived from the ReGAl protocol. Specifically, inclusion criteria for NCI were: (a) age- and education-adjusted MMSE score higher or equal to 24 indicating good general cognitive functions, (b) normal performance in standardized memory tests and cognitive tasks, (c) scores higher than 4 on ADL and IADL indicative of no impaired functional capacity, and (d) CDR equal to 0. This group consisted of 531 cases aged from 60 to 94 years (56.7% women). Inclusion criteria for MCI were: (a) MMSE score higher or equal to 21 indicating preserved general cognitive functions, (b) objective memory deficit, defined as a pathological score (below the normality cut-off) in at least one standardized memory test, with normal performance in the other cognitive tasks, (c) scores of 4 or higher on ADL and IADL indicative of an adequate functional capacity, (d) CDR lower or equal to 1.00, and (e) no dementia (APA, 1994). This group consisted of 182 cases aged from 61 to 91 years (52.7% women). Inclusion criteria for CI subjects were in line with the assessment criteria for dementia (APA, 1994) referring to mild or moderate levels. Specifically, (a) MMSE score higher or equal to 17 indicating unpreserved general cognitive functions, (b) objective memory deficit, defined as a pathological score (below the normality cut-off) in standardized memory test, (c) low scores on ADL and IADL indicative of functional impairment, and (d) CDR higher or equal to 1.00. This group consisted of 529 cases aged from from 61–91 years (60.7% women). Finally, some cases ($N = 63$) were not clearly classifiable into these three categories due to inconsistencies among scores or missing information. Thus, to avoid incorrect classifications, they were excluded from the analyses.

3. Results

IRT analyses were conducted employing the IRTPRO software (Cai et al., 2011). We applied the unidimensional two-parameter (2PL) logistic model, which is the most commonly used IRT model in clinical assessment (for a review see Thomas, 2011).

Preliminarily, the dimensionality of the GDS-15 was tested in each

group (NCI, MCI, and CI) to verify the possibility of using the unidimensional 2PL model for IRT analyses. Indeed, only if there is a common factor running among the items, the item parameters estimated under this model reflect properly the latent trait (Reise et al., 2010). Thus, we looked at the presence of local dependence (LD), —i.e. an excess of covariation among item responses that is not accounted for by a unidimensional model—, using the χ^2LD statistic (Chen and Thissen, 1997). The presence of values of 10 or greater indicates the presence of multiple factors. From these preliminary analyses we observed that none of the LD statistics were greater than 10 in the three groups.

The item fit under the 2PL model was tested computing the S-χ^2 statistics, and item parameters - indicating the location (*b*) and the discrimination ability (*a*) of the items - were estimated by employing the marginal maximum likelihood estimation method with the expectation–maximization algorithm (Bock and Aitkin, 1981) implemented in IRTPRO. Due to the large sample size α was fixed at 0.01. In each group, all the items fitted under the 2PL model. The discrimination values (*a*) ranged from 0.54 to 2.78 (item 10 was the less discriminative one, while item 3 and 7 were the more discriminative ones) and the location values (*b*) ranged from −0.30 to 2.00 logit[1] across the continuum of the latent trait (Table 1). Since the *b* parameter can be interpreted as the "severity" of the symptom described by the item (i.e., higher the level of the trait on which the item is located, higher the severity of the item), the less and the more severe items were item 2 and item 8 were, respectively.

DIF analysis was performed applying the Item Response Theory Likelihood Ratio test approach (IRTLR; Thissen et al., 1988) implemented in IRTPRO software (Cai et al., 2011). Applying IRTLR modeling, the DIF detection procedure is based on a nested model comparison approach. This procedure involves comparing differences in log-likelihoods (distributed as chi-square) associated with nested models. Specifically, the comparisons were made contrasting the NCI group against the other ones (MCI and CI), and MCI against CI. Initial DIF estimates can be obtained by treating each item as a studied item to determine "anchor" items. Once anchor items are selected (to adjust for multiple comparisons and large sample size, α was fixed at 0.01) and used to estimate the trait, we defined the final DIF status of the GDS-15 items through an iterative process of log-likelihood comparison analyses. The first step of the DIF analysis is reported in Table 2. From the first step no items were identified as having DIF when comparing MCI and CI, while when comparing NCI *vs* MCI and CI (NCI was the reference group), items 6 and item 10 showed DIF ($p < 0.01$). Since DIF analysis examines differences in item parameters, for the 2PL model two types of DIF can be detected: uniform DIF, which refers to location parameters, and non-uniform DIF, which refers to discrimination parameters. Item 6 and 10 were identified as studied items for uniform DIF. Then, using all the other items as "anchor" items, we repeated the analysis. During this iterative process the DIF status of item 6 and item 10 did not change.

To assess the magnitude of the DIF at the item level, we employed the Mantel-Haenszel Common Log-Odds Ratio (MH LOR; Camilli and Shepard, 1994; Mantel and Haenszel, 1959). The MH LOR values can be interpreted as follow: (a) < 0.43 are considered indices of a negligible effect, (b) from 0.43 to 0.64 DIF has a moderate effect, and (c) equal 0.64 or higher are considered a large effect (Penfield and Algina, 2006). In addition, negative values indicate a bias against the reference group and positive values indicate a bias against the focal group. The magnitude of DIF (MH LOR = −0.53) was moderate for item 6, which was biased against the focal group. The magnitude of the DIF (MH LOR = 0.66) was large for item 10, which was biased against the reference group.

To display graphically the differential functioning of an item, we can look at the *Item Characteristic Curves* (ICC). The ICC represents the item location parameter, which reflects the level of the trait where there is a 50% change of endorsing the item, and the item discrimination parameter, which describes the slope of the curve. From the visual inspection of the ICCs of item 6 and 10 (Fig. 1), we can clearly see the differences in the item location across groups: item 6 was less severe for the NCI group when compared to the others, and the item 10 was less severe for the MCI and CI groups when compared to the NCI group. Additionally, Fig. 1 shows that DIF was moderate for item 6, whereas it was large for item 10.

Finally, we followed Penfield and Algina (2006) guidelines about Differential Test Functioning (DTF) to assess the impact of DIF items at the test level. When the variance of DIF effects (*tau-squared*) is approximately 0.14, DTF is large. This happens when 25% or more of the items have a moderate or large magnitude of DIF, i.e., if 25% or more items have an absolute value of MH LOR greater than or equal to 0.43. DTF can be considered medium for $0.07 \leq$ *tau-squared* ≤ 0.14 and small for values smaller than 0.07, indicating that less than 10% of items have moderate or large magnitude of DIF. In the current study, the GDS-15 showed a medium DTF (*tau-squared* = 0.12), i.e., about 13% of the items showed moderate or large DIF.

4. Discussion

A relevant challenge to research and clinical practice on geriatric depression is the diagnosis of the disorder in individuals with cognitive impairment. For this reason, it is important to confirm the accuracy of the most commonly used diagnostic instruments for depression when used with patients with impaired cognitive functioning (Watson and Pignone, 2003). In this vein, the current study aimed to test the presence of measurement biases in the GDS-15 as a function of cognitive functioning applying IRT-based DIF analysis. This method allows to establish empirically whether there is a difference in item responses among people with different levels of cognitive functioning or, in other words, if the likelihood of responding affirmatively to an item does not depend exclusively on depression, but also on cognitive functioning.

Differently from Tang et al. (2005), our results suggested that two items of the GDS-15 function differently in cognitively normal and impaired individuals. Not surprisingly, item 10 ("Do you feel you have more problems with memory than most?") was biased against for the former group. As such, a randomly selected cognitively normal individual with a specific level of depression and a randomly selected cognitively impaired individual with the same level of depression have not the same likelihood of endorsing this item. Since MCI and CI old people are more likely to answer positively, this symptom appears to be a less severe indicator of depression for them. Moreover, whereas the difference among groups is not significant (i.e., the item does not show non-uniform DIF), item 10 appears to be scarcely informative in the CI group. Item 6 ("Are you afraid that something bad is going to happen to you?") was the other item with DIF. This item was more likely to be endorsed by NCI when compared to MCI and CI respondents. As such, item 6 describes a more severe symptom for cognitively impaired people when compared to cognitively normal people, probably because they interpret the item in a more general way, not necessarily related to the dysphoric feelings associated with depression.

At the test level, the presence of DIF items might produce a systematic bias on classifications based on cut-off scores (Hidalgo et al., 2015; Jones and Raju, 2000; Stark et al., 2004). In this case, since very few items work differently, the impact on the total score can be considered moderate. Additionally, DIF seems to propagate through to GDS-15 total scores in such a way that it is not expected to produce group biased total measures. Indeed, one item is biased against cognitively normal people and the other one against cognitively impaired people. Thus, two items of the GDS-15 work differently but a

[1] Parameters are expressed on a log-odd scale and the units are called logits. The logit is the logarithm of the *odd*, i.e., the ratio between the probability of answering "yes" and the probability of answering "no".

Table 1
Chi square fit statistics for each item of the 15-item version of the Geriatric Depression Scale (GDS-15) and for each cognitive functioning group.

	Cognitive Functioning											
	NCI				MCI				CI			
Item	*S-χ^2*	*p*	*a*	*b*	*S-χ^2*	*p*	*a*	*b*	*S-χ^2*	*p*	*a*	*b*
1	12.50	0.41	1.59	1.01	11.13	0.13	2.70	0.83	18.18	0.08	2.03	1.06
2	17.45	0.09	1.53	0.04	12.78	0.12	1.31	−0.13	8.94	0.63	1.17	−0.30
3	10.44	0.49	2.10	0.55	15.38	0.05	1.87	0.53	8.41	0.49	2.78	0.45
4	6.81	0.81	1.59	0.45	10.29	0.33	0.93	0.68	5.53	0.90	1.66	0.33
5	5.42	0.91	1.72	0.43	10.68	0.15	1.95	0.40	5.73	0.89	1.97	0.79
6	11.41	0.49	0.77	0.19	10.68	0.22	0.93	0.42	10.73	0.55	1.02	0.72
7	9.98	0.53	1.98	0.49	11.04	0.09	2.52	0.48	4.97	0.89	2.39	0.56
8	15.89	0.25	0.82	1.99	11.72	0.23	1.09	1.54	9.21	0.76	1.00	2.00
9	14.21	0.29	1.18	0.34	7.93	0.54	0.76	0.24	6.22	0.94	0.57	0.20
10	12.69	0.39	0.84	0.85	12.78	0.17	0.93	−0.13	12.43	0.49	0.54	0.11
11	12.29	0.42	1.30	0.85	9.39	0.31	2.10	0.85	6.60	0.92	1.46	1.09
12	12.83	0.31	2.11	0.98	8.42	0.40	1.45	1.05	4.79	0.90	2.21	0.68
13	11.67	0.39	1.57	0.37	14.06	0.08	1.86	0.27	6.78	0.87	1.50	0.37
14	11.85	0.46	1.75	1.02	16.21	0.04	1.17	1.67	10.81	0.46	1.72	1.06
15	7.97	0.85	1.13	1.46	12.57	0.13	1.69	1.20	5.96	0.95	1.12	1.30

Note: Fit was calculate under the 2PL logistic model. Due to the large sample size α was fixed at 0.01. NCI = no cognitive impairment; MCI = mild cognitive impairment; CI = cognitive impairment. 2PL = two-parameter model, *a* = discrimination, *b* = location.

Table 2
Differential Item Functioning analyses of the 15-item version of the Geriatric Depression Scale (GDS-15) across different cognitive functioning groups.

	NCI vs MCI&CI				*MCI vs CI*			
Item	*a*		*b*		*a*		*b*	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
1	3.3	0.07	1.2	0.27	0.7	0.40	0.4	0.51
2	1.4	0.24	3.9	0.05	0.1	0.73	0.9	0.33
3	0.3	0.60	0.4	0.53	2.9	0.09	0.0	0.96
4	1.3	0.25	0.2	0.66	4.7	0.03	0.5	0.50
5	2.9	0.09	4.4	0.04	1.6	0.20	4.9	0.03
6	1.1	0.30	*8.4*	*0.004*	0.1	0.76	2.9	0.09
7	3.2	0.07	0.9	0.34	1.3	0.26	0.0	0.99
8	0.8	0.38	0.5	0.46	0.0	0.84	1.7	0.19
9	6.5	0.01	2.2	0.13	0.4	0.52	0.2	0.66
10	0.4	0.57	*27.7*	*0.0001*	1.7	0.19	0.5	0.47
11	2.2	0.14	5.3	0.02	1.2	0.28	0.0	0.93
12	0.6	0.45	3.7	0.06	2.7	0.10	1.0	0.31
13	0.1	0.75	0.2	0.64	0.6	0.44	0.2	0.69
14	0.9	0.34	1.4	0.24	1.6	0.21	1.3	0.27
15	0.8	0.36	0.0	0.94	1.2	0.27	2.0	0.16

Note: DIF was calculate under the 2PL logistic model. α was fixed at 0.01 due to the large sample size and to adjust for multiple comparisons (significant differences are in italics). DIF = Differential Item Functioning, 2PL = two-parameter model. NCI = no cognitive impairment; MCI = mild cognitive impairment; CI = cognitive impairment, *a* = discrimination, *b* = location.

differential interpretation of scores is not needed and.

The current psychometric findings partially account for the well-known problems of making a differential diagnosis of depression and dementia. Indeed, the underlying neuropathological condition that causes mild cognitive impairment or dementia also causes depressive symptoms (Panza et al., 2010). Moreover, there is still a controversial debate whether depression represents a risk factor or a prodromal feature of dementia (Ownby et al., 2006), or if dementia may be a risk factor for depression due to a psychological reaction to the cognitive and behavioral changes accompanying dementia. Apart from these specific hypotheses, dementia and depression symptoms partially overlap, and item 10 of the GDS-15 is a clear example. For this reason, as a practical recommendation, it might be useful to check for the positive endorsement of this item in MCI and CI patients, and to evaluate its impact on the total score. When the total score rises around the cut-off, the clinicians should take into account other information about the patient and decide to remove or maintain it for the total score computation and interpretation.

Future studies may confirm the current findings and extend beyond the limitations of the present study. First, whereas one of its strengths is the large size of the NCI and CI samples, the relatively small size of the MCI sample might reduce the stability of the current results. Secondly, the results were obtained with Italian-speaking people. Thus, since culture-specific differences might affect item responses, it would be important to test the performance of the scale with different samples. Specifically, DIF analysis should be applied to test the measurement equivalence of the scale across different linguistic versions.

Finally, some indications can be derived from the current findings to reword or exclude items in order to eliminate possible source of different functioning. In this way, it will be possible to develop a shorter measurement instrument, which provides reliable and bias-free information for screening purpose with a minimal loss of measurement precision. In fact, several abbreviated or short form versions of the GDS scales have been proposed (for a detailed review see Pocklington et al., 2016), but many of them include at least one item with DIF. For example, one or both DIF items (i.e., item 6 and item 10) were included in the various 10-item versions of the GDS (Castelo et al., 2010; D'Ath et al., 1994; Izal et al., 2010) as well as in the 4-item versions proposed by D'Ath et al. (1994) and Allgaier et al. (2013). Finally, other brief versions –consisting of 4 items (Almeida and Almeida, 1999), 5 items (Hoyl et al., 1999; Italian version: Rinaldi et al., 2003), and 7 items (Broekman et al., 2011)– were derived from the GDS-15 excluding item 6 and item 10, but their equivalent functioning across different cognitive functioning groups was not tested when the subset of items is used as a stand-alone instrument.

Since a screening tool for the elderly is intended to apply to everyone in the target population, the current IRT-based DIF analyses support the widespread use of the GDS-15 to screen geriatric depression.

Acknowledgments

The authors would like to thank the geriatricians of the Institute of Gerontology and Geriatrics of Perugia and Dr Roberta Majla Valentina Trovato for data imputation.

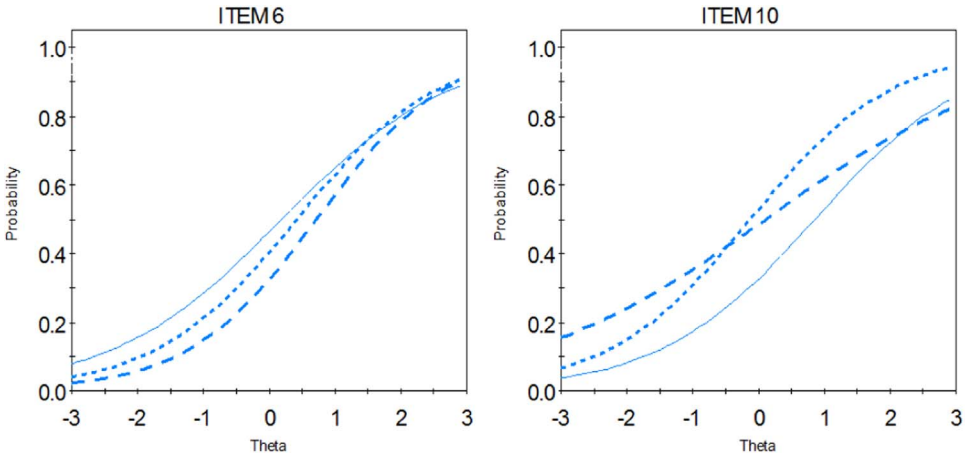

Fig. 1. Item Characteristic Curves of item 6 and item 10 of the 15-item version of the *Geriatric Depression Scale* (GDS-15) for NCI (solid line), MCI (dotted line) and CI (dashed line). Latent trait (Theta) is shown on the horizontal axis, and the probability of item endorsement is shown on the vertical axis. *Note*: NCI = no cognitive impairment; MCI = mild cognitive impairment; CI = cognitive impairment.

Founding source

None.

Disclosure

None.

References

Allgaier, A.K., Kramer, D., Saravo, B., Mergl, R., Fejtkova, S., Hegerl, U., 2013. Beside the Geriatric Depression Scale: the WHO-Five Well-being Index as a valid screening tool for depression in nursing homes. Int. J. Geriatr. Psychiatry 28 (11), 1197–1204.

Almeida, O.P., Almeida, S.A., 1999. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int. J. Geriatr. Psychiatry 14 (10), 858–865.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorder. Author, Washington, DC.

Azulai, A., Walsh, C.A., 2015. Screening for geriatric depression in residential care facilities: a systematic narrative review. J. Gerontol. Soc. Work 58 (1), 20–45.

Berg, L., 1984. Clinical dementia rating. Br. J. Psychiatry 45 (3) (339-339).

Boccardi, V., Baroni, M., Paolacci, L., Ercolani, S., Longo, A., Giordano, M., ReGAl Study Group, 2016. Anticholinergic burden and functional status in older people with cognitive impairment: Results from the ReGAl project. J. Nutr. Health Aging. http://dx.doi.org/10.1007/s12603-016-0787-x.

Bock, R.D., Aitkin, M., 1981. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika 46 (4), 443–459.

Broekman, B.F., Niti, M., Nyunt, M.S.Z., Ko, S.M., Kumar, R., Ng, T.P., 2011. Validation of a brief seven-item response bias-free geriatric depression scale. Am. J. Geriatr. Psychiatry 19 (6), 589–596.

Burke, W.J., Houston, M.J., Boust, S.J., Roccaforte, W.H., 1989. Use of the Geriatric Depression Scale in dementia of the Alzheimer type. J. Am. Geriatr. Soc. 37 (9), 856–860.

Cai, L., Thissen, D., du Toit, S.H.C., 2011. IRTPRO 2.1 for Windows. Scientific Software International, Chicago, IL.

Camilli, G., Shepard, L.A., 1994. Methods for Identifying Biased Test Items. Sage, Newbury Park, CA.

Carlesimo, G.A., Caltagirone, C., Gainotti, G.U.I.D., Fadda, L., Gallassi, R., Lorusso, S., Parnetti, L., 1996. The mental deterioration battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. Eur. Neurol. 36 (6), 378–384.

Castelo, M.S., Coelho-Filho, J.M., Carvalho, A.F., Lima, J.W., Noleto, J.C., Ribeiro, K.G., Siqueira-Neto, J.I., 2010. Validity of the Brazilian version of the Geriatric Depression Scale (GDS) among primary care patients. Int. Psychogeriatr. 22 (01), 109–113.

Chen, W.H., Thissen, D., 1997. Local dependence indexes for item pairs using item response theory. J. Educ. Behav. Stat. 22 (3), 265–289.

Conradsson, M., Rosendahl, E., Littbrand, H., Gustafson, Y., Olofsson, B., Lövheim, H., 2013. Usefulness of the Geriatric depression scale 15-item version among very old people with and without cognitive impairment. Aging Ment. Health 17 (5), 638–645.

D'Ath, P., Katona, P., Mullan, E., Evans, S., Katona, C., 1994. Screening, detection and management of depression in elderly primary care attenders. I: the acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. Fam. Pract. 11 (3), 260–266.

Embretson, S.E., Reise, S.P., 2000. Item Response Theory for Psychologists. Lawrence Erlbaum, Mahwah, NJ.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12 (3), 189–198.

Hidalgo, M.D., Galindo-Garre, F., Gómez-Benito, J., 2015. Differential item functioning and cut-off scores: implications for test score interpretation. Anu. De. Psicol./ UB J. Psychol. 45 (1), 55–69.

Hoyl, M., Alessi, C.A., Harker, J.O., Josephson, K.R., Pietruszka, F.M., Koelfgen, M., Rubenstein, L.Z., 1999. Development and testing of a five-item version of the Geriatric Depression Scale. J. Am. Geriatr. Soc. 47 (7), 873–878.

Izal, M., Montorio, I., Nuevo, R., Perez-Rojo, G., Cabrera, I., 2010. Optimising the diagnosticperformance of the Geriatric Depression Scale. Psychiatry Res. 178 (1), 142–146.

Jones, J.A., Raju, N.S., 2000. Differential item and test functioning and cutoff scores in personnel decision making. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Jongenelis, K., Pot, A.M., Eisses, A.M.H., Gerritsen, D.L., Derksen, M., Beekman, A.T.F., Ribbe, M.W., 2005. Diagnostic accuracy of the original 30-item and shortened versions of the Geriatric Depression Scale in nursing home patients. Int. J. Geriatr. Psychiatry 20 (11), 1067–1074.

Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W., 1963. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. Jama 185 (12), 914–919.

Lach, H.W., Chang, Y.P., Edwards, D., 2010. Can older adults with dementia accurately report depression using brief forms? Reliability and validity of the Geriatric Depression Scale. J. Gerontol. Nurs. 36 (5), 30–37.

Lawton, M.P., Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 9 (3), 179–186.

Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., Riedel-Heller, S.G., 2012. Age-and gender-specific prevalence of depression in latest-life–systematic review and meta-analysis. J. Affect. Disord. 136 (3), 212–221.

Mantel, N., Haenszel, W., 1959. Statistical aspects of the analysis of data from retrospective studies of disease. J. Natl. Cancer Inst. 22, 719–774.

Mariani, E., Monastero, R., Ercolani, S., Rinaldi, P., Mangialasche, F., Costanzi, E., Mecocci, P., 2008. Influence of comorbidity and cognitive status on instrumental activities of daily living in amnestic mild cognitive impairment: results from the ReGAl project. Int. J. Geriatr. Psychiatry 23 (5), 523–530.

McGivney, S.A., Mulvihill, M., Taylor, B., 1994. Validating the GDS depression screen in the nursing home. J. Am. Geriatr. Soc. 42 (5), 490–492.

Mitchell, A.J., Bird, V., Rizzo, M., Meader, N., 2010. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: a meta-analysis of GDS30 and GDS15. J. Affect. Disord. 125 (1–3), 10–17.

Ownby, R.L., Crocco, E., Acevedo, A., John, V., Loewenstein, D., 2006. Depression and risk for Alzheimer disease: systematic review, meta-analysis, and metaregression analysis. Arch. Gen. Psychiatry 63 (5), 530–538.

Panza, F., Frisardi, V., Capurso, C., D'Introno, A., Colacicco, A.M., Imbimbo, B.P., Capurso, A., 2010. Late-life depression, mild cognitive impairment, and dementia: possible continuum? Am. J. Geriatr. Psychiatry 18 (2), 98–116.

Penfield, R.D., Algina, J., 2006. A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. J. Educ. Meas. 43 (4), 295–312.

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. Int. J. Geriatr. Psychiatry 31, 837–857.

Reise, S.P., Moore, T.M., Haviland, M.G., 2010. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. J. Personal. Assess. 92 (6), 544–559.

Reise, S.P., Waller, N.G., 2009. Item response theory and clinical measurement. Annu. Rev. Clin. Psychol. 5, 27–48.

Rinaldi, P., Mecocci, P., Benedetti, C., Ercolani, S., Bregnocchi, M., Menculini, G., Cherubini, A., 2003. Validation of the five-item geriatric depression scale in elderly subjects in three different settings. J. Am. Geriatr. Soc. 51 (5), 694–698.

Smalbrugge, M., Jongenelis, L., Pot, A.M., Beekman, A.T., Eefsting, J.A., 2008. Screening for depression and assessing change in severity of depression. Is the Geriatric Depression scale (30, 15-and 8-item versions) useful for both purposes in nursing home patients? Aging Ment. Health 12 (2), 244–248.

Spinnler, H., Tognoni, G., 1987. Italian group on the neuropsychological study of ageing:

Italian standardization and classification of neuropsychological tests. Ital. J. Neurol. Sci. 6, 1–120.
Stark, S., Chernyshenko, O.S., Drasgow, F., 2004. Examining the effects of differential item (functioning and differential) test functioning on selection decisions: when are statistically significant effects practically important? J. Appl. Psychol. 89 (3), 497.
Tabert, M.H., Albert, S.M., Borukhova-Milov, L., Camacho, Y., Pelton, G., Liu, X., Devanand, D.P., 2002. Functional deficits in patients with mild cognitive impairment Prediction of AD. Neurology 58 (5), 758–764.
Tang, W.K., Wong, E., Chiu, H.F., Lum, C.M., Ungvari, G.S., 2005. The Geriatric Depression Scale should be shortened: results of Rasch analysis. Int. J. Geriatr. Psychiatry 20 (8), 783–789.
Thissen, D., Steinberg, L., Wainer, H., 1988. Use of item response theory in the study of group differences in trace lines. In: Wainer, H., Braun, H.I. (Eds.), Test Validity. Erlbaum, Hillsdale, N.J., pp. 147–169.
Thomas, M.L., 2011. The value of item response theory in clinical assessment: a review. Assessment 18 (3), 291–307.
Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., Friedrich, F., 2006. The criterion validity of the Geriatric Depression Scale: a systematic review. Acta Psychiatr. Scand. 14 (6), 398–410.
Wang, S., Blazer, D.G., 2015. Depression and cognition in the elderly. Annu. Rev. Clin. Psychol. 11, 331–360.
Watson, L.C., Pignone, M.P., 2003. Screening accuracy for late-life depression in primary care: a systematic review. J. Fam. Pract. 52 (12) (956-956).
Yesavage, J.A., Sheikh, J.I., 1986. Geriatric depression scale (GDS) recent evidence and development of a shorter version. Clin. Gerontol. 5 (1–2), 165–173.
Zumbo, B.D., 2003. Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. Lang. Test. 20 (2), 136–147.

Francesca Chiesi is associate professor in Psychometrics. Her expertise and research interests refer to the development and adaptation of tests applying Item Response Theory (IRT), with a particular focus on the psychometric properties of short forms of test for the assessment of cognitive functioning, personality characteristics, and psychological disorders.

Caterina Primi is associate professor in Psychometrics and Research Methodology. Her expertise and research interests include the construction and adaptation of tests, the cognitive processes involved in statistical and probabilistic reasoning, the relationship between mathematical anxiety and mathematical reasoning, and behavioral addictions in adolescents.

Martina Pigliautile is a Psychologist and she's now attending "Psychotherapy Institute of Expressive Gestalt" in Perugia, Italy. She obtained her Ph.D. in “Cognitive Psychology, Psychophysiology and Personality” at the University of Rome “La Sapienza”, Italy. She collaborated with the Institute of Gerontology and Geriatrics, University of Perugia.

Marta Baroni, born in 1982, is medical doctor from 2008 and geriatrician from 2014. She works at hospital of Perugia, her topics are dementia, osteoporosis and orthogeriatrics care. She's working for European project "PredictND" for dementia prevention. She worked in nursing home in 2014 and 2016.

Sara Ercolani graduated as Doctor of Medicine and Surgery from Perugia University Medical School, Italy, in 2000 and completed her residency in Geriatrics at the Institute of Gerontology and Geriatrics, University of Perugia, in 2005. She is Medical Assistant at the Unit of Geriatrics of the Hospital of Perugia since 2011. She was co-author in more than 20 peer-reviewed publications in reputed journals mainly focused on dementia.

Lucia Paolacci, Geriatric Physician, graduated from University of Perugia (Italy). Medical Director in Nursing Home; Team Member of Prof. Patrizia Mecocci’s group for the ReGAl Project 2.0 and co-author in research papers mainly focused on geriatric neuropsychology and dementia.

Dr Virginia Boccardi has completed her MD degree and PhD at the University of Campania Luigi Vanvitelli. During her career, she has spent almost two years as a research fellow at The University of Medicine and Dentistry of New Jersey (USA). She is Assistant Professor at the Insitute of Gerontology and Geriatrics of the University of Perugia since 2015. She has published more than 45 peer-reviewed publications in reputed journals and has been serving as an editorial board member of repute.

Patrizia Mecocci graduated as Doctor of Medicine and Surgery from Perugia University Medical School, Italy. She then obtained her PhD in ‘Biology and Physiopathology of Aging’ at the University of Modena, Italy. She completed her residency in Geriatrics at the Institute of Gerontology and Geriatrics, University of Perugia, becoming full Professor there in 2006. At present she is Director of the Institute of Gerontology and Geriatrics, of the School of Specialization in Geriatrics and of the Course in Speech Therapy at the University of Perugia. During her career, she has spent time as a research fellow at the University of Lund in Sweden and at Harvard Medical School in the USA, and as visiting professor at the Karolinska Institute in Stockholm. She has authored/co-authored more than 260 peer-reviewed publications, and has contributed 20 books/chapters and monographs. She is a regular contributor as a speaker at international congresses.