International Psychogeriatrics (2013), 25:1, 71–81 © International Psychogeriatric Association 2012
doi:10.1017/S1041610212001421

A meta-analysis of the factor structure of the Geriatric Depression Scale (GDS): the effects of language

Giyeon Kim,[1,2] Jamie DeCoster,[3] Chao-Hui Huang[2] and Ami N. Bryant[2]

[1] *Center for Mental Health and Aging, The University of Alabama, Tuscaloosa, Alabama, USA*
[2] *Department of Psychology, The University of Alabama, Tuscaloosa, Alabama, USA*
[3] *Center for Advanced Study of Teaching and Learning, University of Virginia, Charlottesville, Virginia, USA*

ABSTRACT

Background: Given the lack of consensus on the factor structure of the Geriatric Depression Scale (GDS), as well as the fact that the GDS factor structure appears to vary across diverse cultural and/or language groups, the present meta-analysis examined whether the factor structure of the GDS varies by language.

Methods: A total of 26 published studies using exploratory factor analysis (14,669 participants; 10 languages) were included in the meta-analysis. The factor structure of the GDS was assessed in the overall sample as well as in each language that had been examined in at least two different studies.

Results: The analysis of the full sample resulted in a four-factor structure, whereas analyses of the individual languages produced structures with 4 to 6 factors. The mean variable cosines between languages ranged from 0.612 to 0.839, suggesting that the different languages produced distinct factor structures. The three factors of dysphoria, social withdrawal-apathy-cognitive impairment, and positive mood were commonly observed across different languages. Of these, the positive mood factor was the most similar across the languages.

Conclusions: These results provide strong evidence of language differences in the factor structure of the GDS. The findings suggest a need for researchers and clinicians to be careful when administering the GDS in different languages, as well as a need to take structural differences into account when interpreting results of the GDS.

Key words: Geriatric Depression Scale (GDS), meta-analysis, factor analysis, language, factor structure, depressive symptoms

Introduction

Developed almost three decades ago, the Geriatric Depression Scale (GDS; Yesavage and Brink, 1983) has been one of the most commonly used depression screening tools administered in geriatric populations. The GDS is a self-report scale consisting of 30 items (10 worded negatively and 20 worded positively) answered using a yes/no response format. Values on the scale range from 0 to 30, with higher values indicating more symptoms of depression. A score of 10 is typically used as the threshold to separate patients into depressed and non-depressed groups (Brink *et al.*, 1982). To alleviate respondent fatigue, several short forms of the GDS have been developed. Among these, the 15-item version (Sheikh and Yesavage, 1986) is the most frequently used. Additional short forms that have been examined in research include versions with 10 items (D'Ath *et al.*, 1994; van Marwijk *et al.*, 1995), 5 items (Hoyl *et al.*, 1999), 4 items (D'Ath *et al.*, 1994), and 1 item (D'Ath *et al.*, 1994). While the GDS was originally developed in English, it has been translated into more than 30 different languages including Chinese, French, Greek, Japanese, Italian, Turkish, and Spanish (http://www.stanford.edu/~yesavage/GDS.html).

There are several strengths of the GDS over other depression screening tools. One of the important features of the GDS is that the scale was developed specifically for older adults. As most available depression screening tools were developed and validated in samples of medically healthy younger adults, certain depressive symptoms (such as sleep difficulties and decreased energy) included in those screening tools are often found among non-depressed older adults (Lichtenberg, 2010). The GDS, on the other hand, limits itself to symptoms of depression that are not otherwise likely to appear in older adults. The GDS does not contain somatic

Correspondence should be addressed to: Giyeon Kim, Ph.D., Center for Mental Health and Aging and Department of Psychology, The University of Alabama, 207 Osband Hall, Box 870315, Tuscaloosa, AL 35486 USA. Phone: 205-348-7517; Fax: 205-348-7520. Email: Giyeon.Kim@ua.edu. Received 23 Mar 2012; revision requested 5 Jun 2012; revised version received 19 Jun 2012; accepted 25 Jul 2012. First published online 29 August 2012.

symptom items such as loss of appetite and effort (Kessler, *et al.*, 1992). These somatic complaints are likely to be shared by those with medical disorders, older adults, or people from certain cultural groups (Lewis-Fernández *et al.*, 2005; Kalibatseva and Leong, 2011), making them less valid indicators for depression for members of those groups. Finally, the GDS is easier for older adults to complete because its items are written using relatively simple language using an easy, yes/no response format.

A number of psychometric studies have reported good reliability and validity for the GDS (e.g Yesavage and Brink, 1983; Burke *et al.*, 1995). Previous studies also suggest that the reliability of the GDS is satisfactory in diverse elderly populations (Mui *et al.*, 2002; Mui *et al.*, 2003; Malakouti *et al.*, 2006). For example, the GDS has been found to have a Cronbach's alpha of 0.90 among Iranian elders (Malakouti *et al.*, 2006). Mui and colleagues (2003) also found both long and short forms of the GDS to be reliable measures for assessing depression among community-dwelling elderly Asian immigrants.

The factor structure of the GDS has been examined in a number of previous studies (e.g. Parmelee *et al.*, 1989; Sheikh *et al.*, 1991; Abraham *et al.*, 1994; Mui, 1996a; Adams *et al.*, 2004; Lai *et al.*, 2010). The reported number of factors of the GDS has ranged from two (Chiu *et al.*, 1994; Incalzi *et al.*, 2003; Brown *et al.*, 2007) to nine (Salamero and Marcos, 1992), depending upon versions of the GDS, factor analytic methods (exploratory factor analysis vs. confirmatory factor analysis), and study population. Using an exploratory factor analysis (EFA), the factor structure of the GDS was first reported by Sheikh and colleagues (1991) with the following five factors: (1) sad mood and pessimistic outlook (9 items); (2) a lack of mental and physical energy (6 items); (3) positive or happy mood (6 items); (4) agitation or restlessness (3 items); and (5) social withdrawal (2 items). Sheikh and colleagues (1991) found that four items (i.e. #1 "satisfied," #3 "life empty," #14 "memory," and #17 "feel worthless") did not load on any of the factors. Adams and colleagues (2004) conducted a confirmatory factor analysis (CFA) using 26 items of the GDS (excluding four items – #14 "memory," #11 "restless and fidgety," #24 "frequently get upset," #27 "enjoy getting up in the morning"), and confirmed the five-factor structure (dysphoric mood, withdrawal-apathy-vigor, hopelessness, cognitive, and anxiety).

Previous studies found evidence that the factor structure of the GDS may vary depending on respondents' language use and cultural background (e.g. Eartan and Eker, 2000; Jang *et al.*, 2001; Lai *et al.*, 2005; Pocinho *et al.*, 2009). For example, in a study conducted among older Koreans (interviewed in Korean) and Americans (interviewed in English), Jang and colleagues (2001) found few similarities in the factor structures of the GDS-short form (SF; the 15-item version) between the two language/cultural groups. In their study, a three-factor structure of the GDS-SF ("internal perceptions," "external aspects," and a group containing two items of "staying at home" and "problem with memory") was found among Koreans interviewed in Korean, whereas a four-factor structure ("positive mood," "sad mood/pessimistic outlook," a third group containing incongruent items, and a single-item group containing "staying at home") was found among Americans interviewed in English. The authors suggested that the different structures may reflect different interpretations of and responses to depressive symptom items included in the GDS (Jang *et al.*, 2001). In another study conducted among Turkish older adults (interviewed in Turkish), Eartan and Eker (2000) found a seven-factor structure of the 30-item version of the GDS, but argued strongly that a two-factor structure (negatively-worded and positively-worded factors) would be easier for interpretation in this language/cultural group. Taken together, the literature provides evidence that the factor structure of the GDS may vary across different language versions.

Given the lack of consensus on the factor structure of the GDS, as well as the fact that the GDS factor structure appears to vary across diverse cultural and/or language groups, the present study sought to systematically review literature on the factor structure of the GDS using a meta-analytic approach. The main purposes of this meta-analytic study were to (1) determine as to what factor structure fits best the data provided by prior factor analyses and (2) identify the effects of language on the factor structure of the GDS.

Methods

Initial literature search

A two-phase search strategy used in previous research (Kim *et al.*, 2011) was employed to identify articles for the current meta-analysis. The initial search phase focused on identifying the full range of articles that may potentially be related to the topic of interest. Computerized indices were searched where terms related to GDS, such as "*GDS,*" and *"Geriatric Depression Scale"* were paired with terms related to factor analysis, such as "*factor,*" "*component,*" and "*structure.*" These paired terms were submitted to the following online databases to identify articles that had examined

the factor structure of the GDS in different languages: PSYCINFO (May 1980–August 2010), PubMed (May 1980–August 2010), and Digital Dissertations (May 1980–August 2010). We have also ensured our search from the three databases covers all possible studies for inclusion by submitting our search to the Google Scholar search engine. After excluding duplicated articles, a total of 629 potential articles were found.

Inclusion criteria

The second search phase aimed at finding articles that specifically met the study-inclusion criteria. The abstracts of the 629 articles identified in the initial search phase were reviewed, and 41 articles that included an exploratory factor analysis (EFA) of the GDS were identified. Articles reporting confirmatory factor analyses (CFAs) were not included in the sample due to the following reasons: (1) the inclusion of CFAs would allow researchers' expectations to influence results (because CFAs are model-driven; Kim *et al.*, 2011), (2) meta-factor analyses have traditionally focused on EFAs, and (3) only two studies using CFAs were identified in our search. (Please note that the two studies conducted both CFA and EFA, and thus EFA results of the two studies were included in our study.) The 41-factor analysis articles were then reviewed to locate studies where (1) the factor analysis of the GDS was conducted either on the long-form version (30-item/27-item) or on the short-form version (15-item) of the GDS; (2) the factor-analytic results were reported separately by language or the study only used a single language version; (3) the results provided adequate information to determine which items loaded on each factor for each language group; and (4) the sample in the study did not overlap with those from other articles that had already been included. When two studies shared participants, only the one that had larger sample size was screened in so that no redundant samples were included, which would falsely inflate the study population. All four of the authors jointly decided on the inclusion criteria and made decisions jointly whenever there was any question about whether a study met the criteria.

Studies were not restricted with regard to extraction or rotation methods because the goal was to identify the robust structural characteristics of the GDS across a broad range of analytic strategies. A total of 26 studies that examined the GDS in ten different languages (i.e. Chinese, English, Greek, Hindi, Iranian, Italian, Japanese, Korean, Portuguese, and Turkish) met the above-mentioned criteria and were included in the current analysis. Although we would have excluded any studies if their results suggested that the analysis was not performed correctly, we did not see evidence of this in any of the studies we examined. Thus, we did not exclude any studies on a methodological basis.

Data analysis

The current meta-analysis was based on the meta-analytic methods suggested by Loeber and Schmaling (1985) and Frick and colleagues (1993), which have recently been used by Shafer (2006) and Kim and colleagues (2011). The procedure is as follows:

First, the factor loadings were recorded for each study in the literature that performed a factor analysis of the GDS where the entire sample was assessed in the same language. If a study reported multiple factor analyses that had each been performed on the same language, each was recorded separately. Second, the factor loadings were used to create a co-occurrence matrix with 15 rows and 15 columns for each of these analyses, indicating whether each pair of items in the short version of the GDS did or did not load on the same factor. A pair of items was identified as loading on the same factor when they both had their highest loadings on the same factor and the magnitudes of both of those loadings were at least 0.30. Items were not considered to have loaded on the same factor when the loading magnitudes were both greater than 0.30 but one of the items had a higher loading on a different factor. Secondary loadings were not considered when generating the co-occurrence matrices. Studies were included if they used the full version of the GDS, but the co-occurrence matrix only included the results for items that are part of the short version. Third, the co-occurrence matrices were averaged within each language that had more than one representative factor analysis (i.e. Chinese, English, Japanese, and Korean). This average is the "aggregate correspondence matrix," which indicated the proportion of the factor analyses in which each pair of items loaded on the same factor. For example, when determining the aggregate correspondence matrix, if the first item and the second item loaded on the same factor three times out of a total of five analyses, the entry in the first row and second column (as well as the entry in the second row and first column) would be equal to 0.6. These proportions can be thought of as measures of the similarity between each pair of items, ranging from 0 (not at all similar) to 1 (very similar). We also created an overall aggregate correspondence matrix that averaged across all of the studies included in this meta analysis, including those from languages that only had a single representative factor

analysis. A meta-factor analysis was performed by using the full 30-item version of the GDS, but none of the resulting aggregate correspondence matrices was positive definite and so could not be analyzed.[1]

The final step was to perform an EFA on each of the aggregate correspondence matrices to determine the factor structure of the GDS separately for each of these samples. Although factor analyses are typically performed on correlation matrices, the aggregate correspondence matrices also possess the qualities needed to act as the basis of a factor analysis (Tatsuoka, 1971). Principle components extraction was used, keeping any factors with eigenvalues greater than 1. The Scree plots of the eigenvalues were also examined, but did not provide any evidence that basing the factor selection on eigenvalues distorted the results. After determining the number of factors in the solution, a varimax rotation was used on the extracted factors to simplify the interpretation of the results.

Results

Characteristics of studies meeting the meta-analytic inclusion criteria

Table 1 summarizes the descriptive characteristics of the studies included in the meta-analysis divided by language. There were between one and ten studies for each of the ten included languages. One study (Jang *et al.*, 2001) contributed to both the English and Korean language samples. Some studies analyzed more than one subgroup; the number of subgroups included per study ranged from one to three. The overall combined sample size including all language groups was 14,669 with a mean sample size of 1,467 per language. The Portuguese language had the smallest sample size ($n = 200$), whereas the English language had the largest ($n = 5{,}543$). The publication year of the studies ranged from 1989 to 2010. The number of GDS factors found in the included studies ranged from two to nine. The English language had the largest number of studies, subgroups, and within-language factor variability in the number of extracted factors. A summary of basic characteristics of all 26 included studies is displayed in Table 2.

Table 1. Descriptive characteristics of the GDS studies included in the meta-analysis by language ($n = 26$)

	LANGUAGE GROUPS								
CHARACTERISTIC	CHINESE	ENGLISH	GREEK	HINDI	IRANIAN	ITALIAN	JAPANESE	KOREAN	PORTUGUESE	TURKISH
Number of studies	5	10	1	1	1	1	3	3	1	1
Number of subgroups	7	12	2	1	1	1	3	6	1	3
Number of participants	2,384	5,543	271	1,554	204	2,032	906	1,269	200	306
Publication years (range)	1994–2010	1989–2009	1999	1999	2006	2003	2001–2006	2001–2008	2009	2000
Number of GDS factors (range)	2–4	2–9	4	4	3	4	3–4	3–6	3	7

Notes. One article (Jang *et al.*, 2001) contributed to both the English and Korean meta-analyses; While all 26 articles were included in our meta-analysis, only four language groups (Chinese, English, Japanese, and Korean) having two or more studies were separately analyzed.

[1]The fact that the aggregate correspondence matrix was not positive definite means that after averaging, the results produced a set of covariances that would have been impossible to achieve in the real world. This does not necessarily mean that the 30-item version had a poor structure. It simply means that we were unable to test the structure.

Table 2. Summary of the GDS studies included in the meta-analysis ($n = 26$)

CITATION	LANGUAGE	SUBGROUPS	NUMBER OF PARTICIPANTS (n)	AGE RANGE (M ± SD)
1. Abraham *et al.* (1994)	English	(1a) American	917	71–97 (84.4 ± 6.1)
2. Bentz and Hall (2008)	English	(2a) American	158	65–101 (78.3 ± 7.0)
3. Brown *et al.* (2007)	English	(3a) American – non demented	249	43–91 (68.65 ± 13.25)
		(3b) American – demented	357	47–102 (76.02 ± 8.83)
		(3c) American – psychiatric inpatients	140	60–95 (75.61 ± 7.05)
4. Chiu *et al.* (1994)	Chinese	(4a) Chinese	193	60–87 (72.8 ± 6.4)
5. Cornett (2009)	English	(5a) American	282	65–96 (79.11 ± 7.29)
6. Ertan and Eker (2000)	Turkish	(6a) Turkish – living at home	179	60–87 (66.6 ± 5.0)
		(6b) Turkish – living in retirement home	97	62–93 (76.5 ± 7.0)
		(6c) Turkish – with major depression	30	60–89 (67.8 ± 6.0)
7. Fountoulakis *et al.* (1999)	Greek	(7a) Greek – non depressed	168	65+ (79.98 ± 5.77)
		(7b) Greek – depressed	103	65+ (80.20 ± 5.46)
8. Friedman *et al.* (2005)	English	(8a) American	960	65–100 (79.3 ± 7.4)
9. Ganguli *et al.* (1999)	Hindi	(9a) Indian	1,554	57–95 (67.3 ± 6.9)
10. Hupp (1998)	English	(10a) American	655	35–85 (61.45 ± 10.5)
11. Incalzi *et al.* (2003)	Italian	(11a) Italian	2,032	60+ (76.3 ± 8.4)
12. Jang *et al.* (2005)	Korean	(12a) Korean American – low acculturation	99	60–92 (69.8 ± 7.05)
		(12b) Korean American – high acculturation	129	
13. Jang *et al.* (2001)	Korean	(13a) Korean	153	60–89 (65.9 ± 5.0)
	English	(13b) American	459	60–84 (72.4 ± 6.2)
14. Kim *et al.* (2008)[a]	Korean	(14a) Korean – with major depression	61	50+ (73.02 ± 8.03)
		(14b) Korean – with minor depression	45	50+ (75.07 ± 7.70)
		(14c) Korean – normal	782	50+ (75.03 ± 8.10)
15. Lai (2009)	Chinese	(15a) Chinese – Guangzhou	278	65+ (72.5 ± 6.1)
		(15b) Chinese – Hong Kong	304	65+ (72.9 ± 5.8)
		(15c) Chinese –Taipei	309	65+ (73.5 ± 6.0)
16. Lai *et al.* (2005)	Chinese	(16a) Canadian-Chinese	1,136	55–101 (69.77 ± 8.7)
17. Lai *et al.* (2010)	Chinese	(17a) Chinese	114	60+ (79.08 ± 6.5)
18. Malakouti *et al.* (2006)	Iranian	(18a) Iranian	204	59+ (n.a. ± n.a.)
19. Mui (1996b)	Chinese	(19a) Chinese	50	62–91 (75.1 ± 6.5)
20. Onishi *et al.* (2006)	Japanese	(20a) Japanese	607	65+ (79.4 ± 9.6)
21. Onishi *et al.* (2004)	Japanese	(21a) Japanese	198	65+ (77.3 ± 6.8)
22. Parmelee *et al.* (1989)	English	(22a) American	806	61–99 (83.8 ± n.a.)
23. Pocinho *et al.* (2009)	Portuguese	(23a) Portuguese	200	65–92 (76.63 ± 6.44)
24. Salamero and Marcos (1992)	English	(24a) American	234	60–95 (77.5 ± 7.7)
25. Schreiner *et al.* (2001)	Japanese	(25a) Japanese	101	40+ (69.35 ± 10.35)
26. Sheikh *et al.* (1991)	English	(26a) American	326	66–92 (71.0 ± 4.35)

Notes: M = Mean; SD = Standard Deviation; n.a. = not available
[a]Kim *et al.* (2008) only reported overall factor loadings that included all 3 subgroups combined.

Table 3. Factor names of highest factor loadings of the GDS items by language

GDS ITEMS	OVERALL	CHINESE	ENGLISH	JAPANESE	KOREAN
1: Satisfied	P	D	P	P	P
2: Dropped Interests	S	S	S	D	–
3: Empty	D	D	–	D	D
4: Bored	D	D	D	D	D
7: Good Spirits	P	P	P	P	P
8: Afraid	D	D	S	D	D
9: Happy	P	D	P	P	P
10: Helpless	D	P	D	D	n/a
12: Stay Home	S	S	S	S	–
14: Memory Problems	S	S	S	S	–
15: Wonderful	P	n/a	P	P	P
17: Worthless	–	D	D	–	–
21: Energy	P	P	–	–	P
22: Hopeless	–	–	D	–	P
23: Better Off	–	–	D	D	–
Total Number of Factors	4	4	4	5	6

Notes. P = positive mood; S = social withdrawal-apathy-cognitive impairment; D = dysphoria; – = Items did not load on any of the three common factors; n/a = No information available because of a lack of convergence

Language effects on factor structure

The rotated factor loadings from the meta-analysis for each language are presented in Appendix I. These are summarized in Table 3, which reports the factor each item of the GDS loads on for each language. Given that meta-analysis required multiple studies for synthesis, we only meta-analyzed the four language groups (Chinese, English, Japanese, and Korean) that had two or more studies. Other six language groups (Greek, Hindi, Iranian, Italian, Portuguese, and Turkish) having only one study were not separately analyzed, but were included in the overall meta-analysis. The initial aggregate correspondence matrix for the Chinese version of the GDS produced a non-positive definite matrix, and so could not be analyzed. Dropping either one or two items from the matrix was never sufficient to produce a positive definite matrix, but we did find that there were three sets of three items that, when dropped, would lead to an analyzable, positive-definite aggregate correspondence matrix. These sets were (items #7, #15, #17), (items #7, #15, #23), and (items #10, #15, #17). The results for the Chinese version of the GDS presented in Appendix I and referenced in Table B are, therefore, determined by averaging the results across the three analyzable matrices after equating the factors. In addition, the initial aggregate congruence matrix for the Korean version of the GDS was also not positive definite. However, we did obtain a positive-definite matrix when we dropped item 10, so this is what is presented in Appendix I and referenced in Table E.

As shown in Table 3, three common factors appear consistently across most of the languages. The first common factor was primarily composed of "dysphoria" (items #3, #4, #8, and #10), which appeared in all of the solutions. The second common factor was "social withdrawal-apathy-cognitive impairment" (items #2, #12, and #14), which appeared in all of the solutions except that for the Korean language. The third common factor was "positive mood" (items #1, #7, #9, #15, and #21), which appeared in all of the solutions. The remaining factors were more idiosyncratic and did not replicate across languages. Many of these did not appear to have meaningful interpretations, and thus we do not present interpretations of these additional factors.

Global comparisons of the GDS factor structures between languages

After obtaining the factor-analytic results for each language, we determined the extent to which the factor structures were equivalent across the different languages. Paralleling the prior work of Kim and colleagues (2011), we used the method proposed by Kaiser and colleagues (1971) and implemented by Fleming (1992). The Kaiser method was chosen because it is well –established; has an accepted criterion for determining equivalence; and has a software implementation. The Kaiser method allows pair-wise comparisons of the full factor structures between groups to determine the degree to which they are similar. Bushman and colleagues (1991) describe three principal steps to the Kaiser method. First, the variable vectors from both groups must be normalized to ensure that the variables make equal contributions to the final results.

Table 4. Mean variable cosines comparing factor structures between languages

LANGUAGE	OVERALL	CHINESE	ENGLISH	JAPANESE	KOREAN
Overall	–	0.780	0.816	0.820	0.835
Chinese	0.780	–	0.636	0.612	0.725
English	0.816	0.636	–	0.767	0.769
Japanese	0.820	0.612	0.767	–	0.839
Korean	0.835	0.725	0.769	0.839	–

Notes: Only four language groups (Chinese, English, Japanese, and Korean) having two or more studies were separately analyzed. Other six language groups (Greek, Hindi, Iranian, Italian, Portuguese, and Turkish) having only one study were included in the overall meta-analysis.

Second, the factor loadings are used to position the variables from the two samples in a common vector space. This provides a representation of how the variables are related to the factors in each of the two samples. Finally, the factor axes are rotated to provide maximal overlap between the two samples. After the solution is obtained, the cosines between pairs of variable vectors (defined across the factors) can be taken as a measure of the equivalence of the factor structure between the two groups. The cosines range from −1 to 1, with values closer to 1 indicating greater congruency. Kaiser *et al.* (1971) suggest that a mean variable cosine of 0.85 or higher indicated that the fit between the samples was "reasonable," although Barrett (1986) recommends a more conservative cutoff of 0.90. These cutoffs are independent of the number of scale items and the number of factors in the scale.

Table 4 presents the mean variable cosines comparing the observed factor structures of the GDS across the different languages. Comparisons with the Chinese version always excluded item 15 and comparisons with the Korean version always excluded item 10 because these items were not part of the aggregate correspondence matrices for these languages. The mean variable cosines obtained when comparing the structure found in the overall sample to the individual languages range from 0.612 to 0.839 with an average of 0.76. None of the mean variable cosines from comparisons between individual languages exceeded Kaiser *et al.*'s (1971) suggested cutoff (0.85) for "reasonable" fit, suggesting that the different languages produced distinct factor structures.

Comparing individual factors between languages

The three common factors of "dysphoria," "social withdrawal-apathy-cognitive impairment," and "positive mood" appeared in almost every solution. However, the loadings for these factors were not always consistent across the different languages. Therefore, congruence coefficients were calculated to determine whether the loadings of these three factors were individually equivalent across the different languages. Unlike the Kaiser method, which assessed the equivalence of the full factor structure between groups, the congruence coefficients assessed the equivalence of individual pairs of factors. The congruence coefficient can have values between −1 and 1, where values closer to 1 indicate a greater correspondence between the two factor definitions. Mulaik (1972) suggests a rule of thumb wherein congruence coefficients greater than 0.90 indicate that the factor loadings are invariant between two groups.

To simplify the presentation of these comparisons, homogenous subsets are provided in Table 5. The purpose of this analysis is to identify language groups that have equivalent definitions for each of the three factors, where equivalence is indicated by having a congruence coefficient greater than 0.90. Comparisons with the Chinese version always excluded item 15 and comparisons with the Korean version always excluded item 10, because these items were not part of the aggregate correspondence matrices for these languages (please see Appendix I Tables A and C). For the "Dysphoria" factor, the structures are equivalent between the Japanese and the Korean versions of the GDS, but the structures for Chinese and English are unique. For the "Social Withdrawal-Apathy-Cognitive Impairment" factor, each language appears to have its own unique structure. For the "Positive Mood" factor, the English, Japanese, and Korean versions appear to have similar structures, but the factor structure observed in the Chinese version is not equivalent to those found in the other languages.

Discussion

The GDS has been widely used in diverse cultural groups and has been implemented in more than 30 different languages. However, prior to this

Table 5. Factor equivalencies across languages

	GDS FACTORS		
LANGUAGE	DYSPHORIA	SOCIAL WITHDRAWAL-APATHY-COGNITIVE IMPAIRMENT (SAC)	POSITIVE MOOD
Overall	A	A	A
Chinese	B	A	B
English	C	B	A
Japanese	A	C	A
Korean	A	n/a	A

Notes. Within a factor, language groups not sharing any Roman letters have distinct factor loadings (i.e., congruence coefficient less than .90); the 'dysphoria' factor includes items #3, #4, #8, and #10; the 'social withdrawal-apathy-cognitive impairment' factor includes items #2, #12, and #14; and the 'positive mood' includes items #1, #7, #9, #15, and #21; There is no subgroup assigned to the SAC factor for the Korean language because that factor did not appear in the Korean solution.

meta-analysis, there was still a lack of consensus on the overall factor structure of the GDS, and only a few studies had investigated how this varies between languages. The present meta-analysis sought to fill the gap by examining the effect of language on the factor structure of the GDS. Using data from 26 published factor analyses of the GDS across ten languages, we found clear evidence of language differences in the factor structure of the GDS. While a number of empirical studies have suggested the possibility of differences in the factor structure of the GDS across diverse cultural or language groups, none has reviewed existing literature systematically using a meta-analytic approach. To our knowledge, this is the first meta-analysis of the GDS factor structure, which has implications for both research and practice.

The most important and intriguing finding was that different languages produced distinct factor structures of the GDS. This was evidenced by comparisons of the mean variable cosines comparing the overall factor structures, as well as the congruence coefficients comparing the factor loadings of the factors that were found in common. While a four-factor structure of the GDS was observed in the overall sample, the number of factors observed in each language ranged from four to six factors. Empirical studies included in our meta-analysis also reported a wider range of factors, ranging from two to nine in different language groups. The combination of these findings clearly suggests that the factor structure of the GDS varies significantly depending on linguistic and/or cultural factors. This implies that researchers and clinicians need to be careful about interpreting the factor scores of the GDS when it is administered in languages other than English.

Similarities in the GDS factor structure across different languages should also be noted. Three common factors (dysphoria, social withdrawal-apathy-cognitive impairment, and positive mood) appeared in almost every language group, although the factor loadings for the individual items were not always consistent. The positive mood factor had the most in common across different languages, especially for English, Japanese, and Korean. In a recent meta-analysis of the Center for Epidemiologic Studies Depression Scale (CES-D) factor structure, items related to positive mood turned out to be most consistent across five different racial and ethnic groups (Kim *et al.*, 2011). The reasons why the positive mood factor has repeatedly been more consistent across different cultural or language groups are not clear. One possibility is that the idea of positivity may be simple enough that it is defined in a common way across cultures and languages. Alternatively, it could be argued that the items included in this factor are typically all worded in the same direction, whereas the other factors are more likely to have items with reversed wordings. If this is the case, the consistency may be more of a methodological artifact than a theoretically meaningful fact. Further research is needed to elucidate reasons for such findings.

Implications for both research and practice should be discussed. Our results show that the GDS includes some common depressive symptom items shared across different languages, as well as language- and/or culture-specific items and factors that are unique to certain groups. Differences in the factor structures arising from language may suggest cultural differences in the conceptualization and expression of depressive symptoms included in the GDS measure (Kim, 2010). In a previous

meta-factor analysis on the CES-D, the authors suggested that different cultures conceptualize, experience, and express depressive symptoms in different ways, and that there may be no equivalent concepts for depression in certain non-Western cultures (Kim *et al.*, 2011). In clinical settings, clinicians and practitioners should be aware that older adults with diverse cultural and language backgrounds may express their depressive symptoms differently when languages other than English are used to administer the GDS. Thus, as previous research suggested (Mui *et al.*, 2002; Kim *et al.*, 2009), clinicians should also recognize that biases stemming from poor equivalence across different languages and/or cultures may produce false estimates of depressive symptoms, and that some adjustments, such as deletion of culturally or linguistically inappropriate items, may be needed. While researchers are working on the equivalence issues of the GDS, bilingual clinicians and practitioners are well suited to work with geriatric patients from diverse cultural or language groups. At minimum, clinicians should be trained to be more culturally and linguistically competent in clinical settings.

Some study limitations should be noted. First, the current meta-analysis did not include articles using CFA, mainly because CFA is not the ideal methodology due to potential introduction of bias. Second, although most of the GDS studies included in our meta-analysis used older adult samples, we were unable to control for the age range that each study used in their factor analyses. Given that the GDS was developed specifically for the geriatric population, it is not certain if the same factor structures exist when the GDS is applied to other age groups, or if its structure differs among the oldest old. Third, only four languages (Chinese, English, Japanese, and Korean) were examined by individual meta-analyses because other languages did not include multiple studies. Meta-analytic methods are designed to aggregate findings across studies, so might be used by future researchers to examine more languages when additional data become available. Fourth, we were limited to examining the items from the 15-item version of the GDS because the aggregate correspondence matrix from the 30-item version was not positive definite. Future researchers might explore variations in the factor structure of the full version of the scale using either primary studies or a different meta-analytic database. Lastly, the current meta-analysis was unable to control for potential confounding factors such as race/ethnicity, gender, socioeconomic status, and the level of acculturation that might affect the factor structures of the GDS. For example, previous research reported that the structure of other mental health screening tools varied by the level of acculturation and race/ethnicity (Chiriboga *et al.*, 2007; Kim *et al.*, 2011). While our investigation on the effect of language may overlap with testing the effect of race/ethnicity, this is not always true; so the relation between the two effects should be explored. Thus, further meta-analytic research separating the effects of language and culture on the GDS factor structure is clearly needed.

In sum, the factor structure of the GDS varied by language. As was suggested in previous research (Kim, 2010; Kim *et al.*, 2011), different factor structures across different cultural and language groups may suggest potential differences in the conceptualization of overall depressive symptoms. When factor analyses for the GDS are conducted, researchers should recognize potential structural differences in the GDS arising from language in which the measure is assessed. Future research should further examine the factor structure of the GDS to confirm the structure. We recommend that researchers use exploratory factor analysis to verify the factor structure of the GDS whenever it is translated into a new language so that structural differences can be taken into account when researchers interpret this measure.

Conflict of interest declaration

None.

Description of authors' roles

Giyeon Kim directed and designed the study, supervised the data collection and analyses, and wrote the paper. Jamie DeCoster was responsible for the statistical design of the study and for carrying out the statistical analyses and contributed to drafting the Methods and Results sections. Chao-Hui Huang searched and collected data for the meta-analysis and assisted with manuscript preparation. Ami N. Bryant searched and collected data for the meta-analysis and assisted with manuscript preparation.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

*__Abraham, I. L., Wofford, A. B., Lichtenberg, P. A. and Holroyd, S.__ (1994). Factor structure of the Geriatric Depression Scale in a cohort of depressed nursing home residents. *International Journal of Geriatric Psychiatry*, 9, 611–617.

Adams, K. B., Matto, H. C. and Sanders, S. (2004). Confirmatory factor analysis of the Geriatric Depression Scale. *The Gerontologist*, 44, 818–826.

Barrett, P. (1986). Factor comparison: an examination of three methods. *Personality and Individual Differences*, 7, 327–340.

*__Bentz, B. G. and Hall, J. R.__ (2008). Assessment of depression in a geriatric inpatient cohort: a comparison of the BDI and GDS. *International Journal of Clinical and Health Psychology*, 8, 93–104.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P., Adey, M. B. and Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist*, 1, 37–44.

*__Brown, P. J., Woods, C. M. and Storandt, M.__ (2007). Model stability of the 15-item Geriatric Depression Scale across cognitive impairment and severe depression. *Psychology and Aging*, 22, 372–379.

Burke, W. J., Roccaforte, W. H., Wengel, S. P., Conley, D. M. and Potter, J. F. (1995). The reliability and validity of the Geriatric Depression Rating Scale administered by telephone. *Journal of American Geriatric Society*, 43(6), 674–679.

Bushman, B. J., Cooper, H. M. and Lemke, K. M. (1991). Meta-analysis of factor analyses: an illustration using the Buss-Durkee Hostility Inventory. *Personality and Social Psychology Bulletin*, 17, 344–349.

Chiriboga, D. A., Jang, Y., Banks, S. and Kim, G. (2007). Acculturation and its effect on symptom structure in a sample of Mexican American elders. *Hispanic Journal of Behavioral Sciences*, 29, 83–100.

*__Chiu, H. F. K, Lee, H, C. B., Wing, Y. K., Kwong, P. K., Leung, C. M. and Chung, D. W. S.__ (1994). Reliability, validity, and structure of the Chinese Geriatric Depression Scale in a Hong Kong context: a preliminary report. *Singapore Medical Journal*, 35, 477–480.

*__Cornett, P. F.__ (2009). *Factors of the Geriatric Depression Scale that may distinguish between four cognitive diagnostic groups: Normal, mild cognitive impairment, dementia of the Alzheimer's type, and vascular dementia.* Available from ProQuest Dissertations and Theses database. (UMI No. 3399404)

D'Ath, P., Katona, P., Mullan, E., Evans, S. and Katona, C. (1994). Screening, detection and management of depression in elderly primary care attenders: the acceptability and performance of the 15-item Geriatric Depression Scale and the development of shorter versions. *Family Practice*, 11, 260–266.

*__Ertan, T. and Eker, E.__ (2000). Reliability, validity, and factor structure of the Geriatric Depression Scale in Turkish elderly: are there different factor structures for different cultures? *International Psychogeriatrics*, 12, 163–172.

Fleming, J. S. (1992). FACTREL: a general program for relating factors between studies using the Kaiser-Hunka-Bianchini method. *Educational and Psychological Measurement*, 52, 113–115. doi:10.1177/0013164492052001015.

*__Fountoulakis, K. N. *et al.*__ (1999). The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. *Aging Clinical and Experimental Research*, 11, 367–372.

Frick, P. J. *et al.* (1993). Oppositional defiant disorder and conduct disorder: a meta-analytic review of factor analyses and cross-validation in a clinical sample. *Clinical Psychology Review*, 13, 319–340.

*__Friedman, B., Heisel, M. J. and Delavan, R. L.__ (2005) Psychometric properties of the 15-item Geriatric Depression Scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of American Geriatric Society*, 53, 1570–1576.

*__Ganguli, M., Dube, S., Johnston, J. M., Pandav, R., Chandra, V. and Dodge, H. H.__ (1999). Depressive symptoms, cognitive impairment and functional impairment in a rural elderly population in India: a Hindi version of the Geriatric Depression Scale (GDS-H). *International Journal of Geriatric Psychiatry*, 14, 807–820.

Hoyl, M. T. *et al.* (1999). Development and testing of a five-item version of the geriatric depression scale. *Journal of the American Geriatrics Society*, 47, 873–878.

*__Hupp, G. S.__ (1998) *Exploratory factor analysis of the Geriatric Depression Scale among cardiac patients.* Available from ProQuest Dissertations and Theses database. (UMI No. 1389489)

*__Incalzi, R. A., Cesari, M., Pedone, C. and Carbonin, P. U.__ (2003). Construct validity of the 15-item Geriatric Depression Scale in older medical inpatients. *Journal of Geriatric Psychiatry Neurology*, 16, 23–28.

*__Jang, Y., Kim, G. and Chiriboga, D.__ (2005). Acculturation and manifestation of depressive symptoms among Korean-American older adults. *Aging and Mental Health*, 9(6), 500–507.

*__Jang, Y., Small, B. J. and Haley, W. E.__ (2001). Cross-cultural comparability of the Geriatric Depression Scale: comparison between older Koreans and older Americans. *Aging and Mental Health*, 5, 31–37.

Kaiser, H. F., Hunka, S. and Bianchini, J. (1971). Relating factor studies based upon different individuals. *Multivariate Behavioral Research*, 6, 409–422. doi:10.1207/s15327906mbr0604_3.

Kalibatseva, Z. and Leong, F. T. (2011). Depression among Asian Americans: review and recommendations. *Depression Research and Treatment, 2011.* doi:10.1155/2011/320902.

Kessler, R. C., Foster, C., Webster, P. S. and House, J. S. (1992). The relationship between age and depressive symptoms in two national surveys. *Psychology and Aging*, 7, 119–126.

Kim, G. (2010). Measuring depression in a multicultural society: conceptual issues and research recommendations. *Hallym International Journal of Aging*, 12, 27–46.

Kim, G., Chiriboga, D. A. and Jang, Y. (2009). Cultural equivalence in depressive symptoms in older White, Black, and Mexican American adults. *Journal of the American Geriatrics Society*, 57, 790–796.

Kim, G., DeCoster, J., Huang, C.-H. and Chiriboga, D. A. (2011). Race/ethnicity and the factor structure of the Center for Epidemiologic Studies Depression Scale: a meta-analysis. *Cultural Diversity and Ethnic Minority Psychology*, 17, 381–396.

*__Kim, J. Y. *et al.*__ (2008). Standardization of the Korean version of the Geriatric Depression Scale: reliability, validity, and factor structure. *Psychiatry Investigation*, 5, 232–238.

*__Lai, D. W. L.__ (2009). Depressive symptoms of elderly Chinese in Guangzhou, Hong Kong, and Taipei. *Aging and Mental Health*, 13, 725–735.

*__Lai, D. W. L., Fung, T. S. and Yuen, C. T. Y.__ (2005). The factor structure of a Chinese version of the Geriatric Depression Scale. *International Journal of Psychiatry in Medicine*, 35, 137–148.

*__Lai, D., Tong, H., Zeng, Q. and Xu, W.__ (2010). The factor structure of a Chinese Geriatric Depression Scale-SF: use with alone elderly Chinese in Shanghai, China. *International Journal of Geriatric Psychiatry*, 25, 503–510.

Lewis-Fernández, R., Das, A. K., Alfonso, C., Weissman, M. M. and Olfson, M. (2005). Depression in US Hispanics: diagnostic and management considerations in family practice. *Journal of American Board of Family Medicine*, 18, 4282–4296.

Lichtenberg, P. A. (ed.). (2010). *Handbook of Assessment in Clinical Gerontology*, 2nd edn. Burlington, MA: Elsevier, Inc.

Loeber, R. and Schmaling, K. B. (1985). Empirical evidence for overt and covert patterns of antisocial conduct problems: a meta-analysis. *Journal of Abnormal Child Psychology*, 13, 337–353.

*__Malakouti, S. K., Fatollahi, P., Mirabzadeh, A., Salavati, M. and Zandi, T.__ (2006). Reliability, validity and factor structure of the GDS-15 in Iranian elderly. *International Journal of Geriatric Psychiatry*, 21(6), 588–593.

Mui, A. C. (1996a). Depression among elderly Chinese immigrants: an exploratory study. *Social Work*, 41, 633–645.

*__Mui, A. C.__ (1996b). Geriatric Depression Scale as a community screening instrument for elderly Chinese immigrants. *International Psychogeriatrics*, 8, 445–458.

Mui, A. C., Burnette, D. and Chen, L. M. (2002). Cross-cultural assessment of geriatric depression: a review of the CES-D and GDS. In J. H. Skinner, J. A. Teresi, D. Holmes, S. M. Stahl and A. L. Stewart (eds.), *Multicultural Measurement in Older Populations* (pp. 147–177). New York: Springer.

Mui, A. C., Kang, S.-Y., Chen, L. M. and Omanski, M. D. (2003). Reliability of the Geriatric Depression Scale for use among elderly Asian immigrants in the USA. *International Psychogeriatrics*, 15, 253–271.

Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York, NY: McGraw-Hill.

*__Onishi, J., Suzuki, Y., Umegaki, H., Endo, H., Kawamura, T. and Iguchi, A.__ (2006). A comparison of depressive mood of older adults in a community, nursing homes, and a geriatric hospital: factor analysis of Geriatric Depression Scale. *Journal of Geriatric Psychiatry and Neurology*, 19, 26–31.

*__Onishi, J., Umegaki, H., Suzuki, Y., Uemura, K., Kuzuya, M. and Iguchi, A.__ (2004). The relationship between functional disability and depressive mood in Japanese older adult inpatients. *Journal of Geriatric Psychiatry and Neurology*, 17, 93–98.

*__Parmelee, P. A., Lawton, M. P. and Katz, I. R.__ (1989). Psychometric properties of the Geriatric Depression Scale among the institutionalized aged. *Psychological Assessment*, 1, 331–338.

*__Pocinho, M. T. S., Farate, C., Dias, C. A., Lee, T. T. and Yesavage, J. A.__ (2009). Clinical and psychometric validation of the Geriatric Depression Scale (GDS) for Portuguese elders. *Clinical Gerontologist: The Journal of Aging and Mental Health*, 32, 223–236.

*__Salamero, M. and Marcos, T.__ (1992). Factor study of the Geriatric Depression Scale. *Acta Psychiatrica Scandinavica*, 86, 283–286.

*__Schreiner, A., Morimoto, T. and Asano, H.__ (2001). Depressive symptoms among poststroke patients in Japan: frequency distribution and factor structure of the GDS. *International Journal of Geriatric Psychiatry*, 16, 941–949.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*, 62, 123–146.

Sheikh, J. I. and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist*, 5, 165–173.

*__Sheikh, J. I. *et al.*__ (1991). Proposed factor structure of the Geriatric Depression Scale. *International Psychogeriatrics*, 3, 23–28.

Tatsuoka, M. (1971). *Multivariate Analysis: Techniques for Educational and Psychological Research*. New York, NY: Wiley.

van Marwijk, H. W. J., Wallace, P, de Bock, G. H., Hermans, J., Kaptein, A. A. and Mulder, J. D.(1995). Evaluation of the feasibility, reliability, and diagnostic value of shortened versions of the geriatric depression scale. *British Journal of General Practice*, 45, 195–199.

Yesavage, J. A. and Brink, T. L. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, 17, 37–49.