

Reliability and Validity of the Center for Epidemiologic Studies Depression Scale (CES-D) in Patients with Systemic Sclerosis

Brett D. Thombs, PhD¹; Marie Hudson, MD,MPH²; Orit Schieir, BA³; Suzanne S. Taillefer, PhD²; Canadian Scleroderma Research Group (CSRG)⁴; Murray Baron, MD²

¹Department of Psychiatry and ²Division of Rheumatology, Sir Mortimer B. Davis – Jewish General Hospital and McGill University, Montreal, Quebec; ³ Department of Epidemiology, Biostatistics and Occupational Health, McGill University; ⁴**CSRG Investigators:** J. Markland, Saskatoon, Saskatchewan; J. Pope, London, Ontario; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; P. Docherty, Moncton, New Brunswick; M. Abu-Hakima, Calgary, Alberta; N. A. Khalidi, Hamilton, Ontario; S. Le Clercq, Calgary, Alberta; E. Sutton, Halifax, Nova Scotia; C. D. Smith, Ottawa, Ontario; E. Kaminska, Hamilton, Ontario; J.-P. Mathieu, Montreal, Quebec; P. Rahman, St. John's, Newfoundland; S. Ligier, Montreal, Quebec.

Grant support: Dr. Baron is the director and Dr. Taillefer is the research coordinator of the Canadian Scleroderma Research Group, which receives grant funding from the Canadian Institutes of Health Research, the Cure Scleroderma Foundation, the Scleroderma Society of Canada, the Ontario Arthritis Society, Actelion Pharmaceuticals, and Pfizer Pharmaceuticals.

Word Count: 2,344

Address for Correspondence:

Brett D. Thombs, Ph.D.

SMBD-Jewish General Hospital

Institute of Community and Family Psychiatry

4333 Cote Ste Catherine Road

Montreal, Quebec H3T 1E4

Tel (514) 340-8222 ext. 5112

E-mail: brett.thombs@mcgill.ca

ABSTRACT

Objective: Reported rates of depressive symptoms in patients with systemic sclerosis (SSc) are high. No depression assessment tools, however, have been validated for SSc patients. The objective of this study was to assess the reliability and construct validity of the Center for Epidemiological Studies Depression Scale (CES-D) in patients with systemic sclerosis (SSc).

Methods: Cross-sectional, multi-center study of 403 SSc patients. Internal consistency reliability was assessed with Cronbach's alpha and structural/construct validity with confirmatory factor analysis (CFA).

Results: Internal consistency reliability was good for the overall CES-D scale ($\alpha=0.88$) and for its 4 factors ($\alpha=0.67$ to 0.88). The 4-factor model originally found in the general population and validated for rheumatoid arthritis patients (Depressed Affect, Somatic/Vegetative, (Lack of) Positive Affect, and Interpersonal factors) fit the data well, as did a second-order version of the same model with an overarching Depression factor that loaded onto each of the 4 first-order factors. The 4-factor model fit the SSc data better than alternative models.

Conclusion: Internal consistency reliability was good, the 4-factor structure reported in the general population was replicated, and a second-order model with an overarching Depression factor fit well. These findings indicate that the CES-D is a valid and reliable measure of depressive symptoms for patients with SSc.

Systemic Sclerosis (SSc), or scleroderma, is a chronic, multi-system disorder of connective tissue characterized by thickening and fibrosis of the skin, and by involvement of internal organs. Patients with SSc report high levels of pain, fatigue, and disability (1). A recent systematic review found that between 36% and 65% of patients with SSc have clinically significant symptoms of depression, a high rate even compared to patients with other acute and chronic conditions (e.g. post-myocardial infarction, congestive heart failure, diabetes, chronic obstructive pulmonary disease, rheumatoid arthritis) when the same assessment tools and scoring cutoffs are used (2). Studies included in that review reported multivariable associations between depressive symptoms and education, overall disease severity, gastrointestinal symptoms, pain, disability, and body image distress, although methodological issues limited the ability to draw strong conclusions about predictors (2). A recent study of 403 SSc patients (of the 470 included in the present study) found that patients with less education, patients who were not married, patients with higher physician-rated overall disease severity, and patients with more tender joints, more gastrointestinal (GI) symptoms, and more difficulty breathing had significantly higher symptoms of depression as measured by the Center for Epidemiologic Studies Depression Scale (CES-D). Patient sex, duration since onset of non-Raynaud's symptoms or since diagnosis, and total skin score or diffuse/limited classification were not significantly associated with depressive symptoms in multivariate analysis (3).

The Center for Epidemiologic Studies Depression Scale (CES-D) is a widely used 20-item self-report measure that was originally designed for assessing depressive symptomatology in the general population (4) (Table 1). It is also commonly used as a general depression screening tool (5). Research supports the CES-D as a valid measure of depressive symptoms among patients with rheumatoid arthritis (RA) (6-8), and Rhee et al. (6) found that the originally

specified CES-D 4-factor model (4) fit RA data well. The 4 factors were Depressive Affect symptoms (7 items), Somatic/Vegetative symptoms (7 items), Interpersonal symptoms (2 items), and (Lack of) Positive Affect symptoms (4 items) (Figure 1). The findings from Rhee et al. are consistent with the results of a recent systematic review (9) that also found strong evidence for the 4-factor model across many different patient groups. This is important because consistency of factor structures across groups provides evidence for construct validity.

No measures of depressive symptoms have been validated for patients with SSc. The objective of this study was to use a large SSc patient sample from a pan-Canadian registry to investigate the internal consistency reliability, convergent validity, and structural/construct validity of the CES-D in patients with SSc.

PATIENTS AND METHODS

Patient Sample. The study sample consisted of patients enrolled in the Canadian Scleroderma Research Group Registry from September 2004 through August 2006 who completed the CES-D. Patients in the Registry were recruited from 15 centers across Canada. To be eligible for the Registry, patients must have a diagnosis of SSc made by the referring rheumatologist, be ≥ 18 years of age, and be fluent in English or French. Registry patients undergo extensive clinical history, physical evaluation, and laboratory investigations and complete a series of self-report questionnaires. Although eventually all SSc patients receiving care from participating centers will be enrolled, for reasons related to resources, this is occurring over time. Thus, this is a convenience, rather than a consecutive, sample. Of patients approached to participate in the CSRG Registry, approximately 90% have enrolled. Patients from all sites provided informed consent, and the research ethics board of each study site approved the data collection protocol.

Measures. The CES-D (4) is a 20-item measure designed to assess the presence and severity of depressive symptomatology. The frequency of occurrence of each symptom during the past week is rated on a 0-3 Likert-type scale (*rarely or none of the time to most or all of the time*), and total scores range from 0 to 60. Standard cutoffs are ≥ 16 for “possible depression” and ≥ 23 for “probable depression” (4). In addition to depressive symptoms, demographic and medical data were collected. Demographic information was based on self-report, and included age, gender, education, marital status, and race/ethnicity. Patients’ medical histories and disease characteristics were obtained via clinical histories and examinations by study physicians. Skin involvement was assessed using the modified Rodnan skin score ranging from 0 to 51 (10). Limited skin disease was defined as skin involvement distal to the elbows and knees with or without face involvement. Disease severity was assessed with a scale developed by Medsger (11), and a severity score of 0 (normal) to 4 (end-stage) was generated for each of the 9 systems.

Self-report measures of mental health function (SF-36 Mental Composite Score [MCS]), physical function (SF-36 Physical Composite Score [PCS]), disability (Health Assessment Questionnaire – Disability Index [HAQ-DI]), and pain (McGill Pain Questionnaire – Short Form [MPQ]) were used to establish convergent validity. Higher scores on the HAQ-DI indicate greater disability, and higher scores on the MPQ indicate greater pain. Both would be expected to be positively associated with higher scores on the CESD. Higher scores on the MCS and PCS indicate better function, and would be expected to be negatively associated with CESD scores. The association between the CESD and the MCS would be expected to be the most robust since the MCS measures mental health and has a strong depression component.

Data Analyses. Internal consistency reliability was evaluated using Cronbach’s alpha. Convergent validity of the CES-D with other self-report measures was assessed using Spearman

correlation coefficients. Confirmatory factor analysis (CFA) models were used to evaluate the factor structure of the CES-D and were conducted with Mplus (version 3.11), explicitly modeling the CES-D items as ordinal data. To do this, Mplus initially estimates item thresholds for ordinal outcome variables using maximum likelihood methods. These estimates are then used to estimate a polychoric correlation matrix. Model parameters are subsequently estimated with weighted least squares using the inverse of the asymptotic covariance matrix as the weight matrix.

Following the methodology that Rhee et al. used with RA patient data (6), six alternative models were compared: [1] a single depression factor model, [2] a 2-factor model of general depression and positive affect, [3a] a 3-factor model combining the Depressive Affect and Somatic/Vegetative factors, [3b] a second 3-factor model combining the Depressive Affect and Positive Affect factors, [4a] Radloff's 4-factor model (Figure 1), and [4b] Radloff's 4-factor model with a second-order depression factor as done by Rhee et al. (6). Second-order factors are global factors composed of all of the first-order factors (e.g., Depressed Affect, Somatic/Vegetative, (Lack of) Positive Affect, and Interpersonal) that provide a mechanism to test the plausibility that a single overarching construct is being measured. All item-factor allocations for each model are shown in Table 1. Although Sheehan et al. (8) reported a slightly different item allocation than Radloff, these models were not tested since they did not fit as well as the Radloff allocations when directly compared with CFA by Rhee et al. (6). Furthermore, the Sheehan item allocations appear to have come from initial exploratory analyses (6) and are not easily justified theoretically.

To assess the fit of the models to the data, practical fit indices were emphasized since chi-square tests of fit are highly sensitive to sample size and can lead to the rejection of well-fitting

models (12). Four practical fit indices were used to evaluate model fit: the Tucker-Lewis Index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). Guidelines proposed by Hu and Bentler (13) suggest that models with TLI and CFI close to .95 or higher, RMSEA close to .06 or lower and SRMR close to .08 or lower are representative of good fitting models.

RESULTS

Sample Characteristics. A total of 470 patients were included in the study. Approximately 86% (n = 397) of the sample was female and approximately 87% was White (n = 408), which is consistent with North American samples from previous reports (14). The mean age of the sample was 55.4 ± 12.6 years, 46% (n = 209 of 457 with data) of patients completed some post-secondary education, and 72% (n = 332) were married or living as married.

The mean duration since onset of non-Raynaud's symptoms was 10.6 ± 8.7 years (median = 8.3), and the mean duration since diagnosis of SSc was 8.4 ± 7.7 years (median = 6.2). Approximately 60% (n = 279 of 464 with data) of patients had diffuse SSc, and mean total skin score was 11.2 ± 10.2 (median = 8.0). Mean disease severity scores for each of the 9 systems were: General = 0.9 ± 1.2 ; Peripheral Vascular = 1.6 ± 1.2 ; Skin = 1.3 ± 0.7 ; Joint/Tendon = 0.9 ± 1.3 ; Muscle = 0.3 ± 0.8 ; Heart = 0.5 ± 1.0 ; Kidney = 0.2 ± 0.7 ; Lung = 1.4 ± 0.7 .

The mean CES-D score was 14.3 ± 10.4 (median = 13.0). Over a third scored at least 16 on the CES-D (n = 178, 37.8%), a standard cutoff for "possible depression," and 20.2% (n = 95) scored ≥ 23 for "probable depression."

Reliability of the CES-D. Overall scale reliability was good ($\alpha = 0.88$) and similar to the values reported in the original validation study ($\alpha = 0.88$ to 0.90) (4). Corrected item-total

correlations for individual items ranged from 0.24 (Item 4, *good*) to 0.73 (Item 6, *depressed*).

Coefficient alphas were also very good for each of the 4 CES-D factors (4): Depressive Affect = 0.88, Somatic/Vegetative = 0.80, Interpersonal symptoms = 0.67, and (Lack of) Positive Affect = 0.82.

Confirmatory Factor Analysis of the CES-D. Factor loadings for all models are shown in Table 1. As shown in Table 2, model fit was best for the original 4-factor model specified by Radloff and for a second-order version of the 4-factor model, which specified a second-order depression factor rather than intercorrelations between the 4 factors. Intercorrelations between the 4 factors in Model 4a ranged from 0.28 to 0.89 and were lowest for correlations that included the Positive Affect factor. In the second-order 4-factor model (Model 4b), the overarching 2nd order Depression factor accounted for 97% of the variance in the Depressed Affect factor (standardized regression coefficient = 0.99), 73% of the variance in the Somatic/Vegetative factor (standardized regression coefficient = 0.86), and 57% of the variance in the Interpersonal Factor (standardized regression coefficient = 0.76), but only 14% of the variance in the Positive Affect factor (standardized regression coefficient = 0.37). More detailed results for all models are available upon request from the corresponding author.

Convergent Validity. Spearman correlations between the CES-D total score and related self-report measures were: MCS = -0.73; PCS = -0.36; HAQ-DI = 0.41; MPQ = 0.44. All correlations were in the expected direction and, as expected, the correlation with the MCS was strong whereas the others were in the moderate range.

DISCUSSION

This study evaluated the reliability and construct validity of the CES-D in a pan-Canadian sample of 470 patients with SSc. The main findings of this study were that the CES-D

had good internal consistency reliability and convergent validity among patients with SSc and that both 1st and 2nd order versions of the standard 4-factor model of the CES-D fit the data well and better than alternative models. Cronbach's alpha for the overall CES-D was 0.88. A widely used standard suggests that self-report measures should have internal consistency reliability of 0.70 or higher and 0.80 or higher for use as a screening tool (15). Given that coefficient alpha is influenced by the number of items in a scale, the internal consistency reliabilities of the CES-D factor subscales were also very strong (0.67 for the 2-item Interpersonal factor to 0.88 for the 7-item Depressed Affect factor). The CFA analysis showed that, consistent with other studies, the CES-D items clustered into 4 interpretable factors: Depressed Affect, Somatic/Vegetative, Interpersonal, and (Lack of) Positive Affect. The good fit of the second-order model supports the use of a total score of the CES-D as a global indicator of levels of depressive symptoms in patients with SSc. Thus, although the Positive Affect scale, for instance, appears to be only weakly related to the overall depressive construct, the total score is a valid measure. Nonetheless, it is possible that a shorter version of the CES-D could provide equally good measurement in SSc, and this should be tested in future work. A strength of this study was that the factor structure was tested with rigorous methods that explicitly modeled the CES-D items as ordinal data.

One limitation of this study is that it did not address criterion-related validity by comparing cutoff scores on the CES-D to a gold standard, such as a structured interview for major depression. Thus, this report establishes that the CES-D is a valid continuous measure of depressive symptoms, but standard cutoff scores for detecting depression need to be verified for SSc patients. One study reported that a cutoff of 19 or higher is best in patients with RA (7), rather than a standard cutoff of 16, although this finding has not been replicated. In addition,

more research is needed to assess the degree to which somatic symptom overlap may bias symptom severity estimates made with the CES-D in patients with SSc, if any. Two studies of patients with RA have reported that several somatic items of the CES-D reflect both depressive symptoms and RA disease factors, although both of these studies concluded that the effect on the CES-D total score was minimal (16, 17). In addition, although recruitment rates were high, this is technically a convenience sample, and characteristics of patients not yet enrolled in the CSRG are not available.

In a recent Delphi exercise (18), the CES-D was proposed by members of the Scleroderma Clinical Trials Consortium as a possible outcome measure in SSc. However, it was found to lack proper validation in this patient population. This study shows that the CES-D is a reliable and valid instrument for measuring depressive symptoms in patients with SSc, although criterion-related validity and specific cutoff scores need to be established.

REFERENCES

1. Haythornthwaite JA, Heinberg LJ, McGuire L. Psychologic factors in scleroderma. *Rheum Dis Clin North Am* 2003;29:427-39.
2. Thombs BD, Taillefer SS, Hudson M, Baron M. Depression in patients with systemic sclerosis: A systematic review of the evidence. *Arthritis Care Res* 2007;57:1089-1097.
3. Thombs BD, Hudson M, Taillefer SS, Canadian Scleroderma Research Group (CSRG), Baron M. Prevalence and clinical correlates of symptoms of depression in patients with systemic sclerosis. *Arthritis Care Res* (in press).
4. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;1:385-401.
5. Williams JW, Jr, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002;24:225-37.
6. Rhee SH, Petroski GF, Parker JC, Smarr KL, Wright GE, Multon KD, et al. A confirmatory factor analysis of the Center for Epidemiologic Studies Depression Scale in rheumatoid arthritis patients: additional evidence for a four-factor model. *Arthritis Care Res* 1999;12:392-400.

7. Martens MP, Parker JC, Smarr KL, Hewett JE, Slaughter JR, Walker SE. Assessment of depression in rheumatoid arthritis: a modified version of the Center for Epidemiologic Studies Depression Scale. *Arthritis Rheum* 2003;49:549-55.
8. Sheehan TJ, Fifield J, Reisine S, Tennen H. The measurement structure of the Center for Epidemiologic Studies Depression Scale. *J Pers Assess* 1995;64:507-21.
9. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol* 2006;62:123-46.
10. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995;22:1281-5.
11. Medsger TA, Jr., Bombardieri S, Czirjak L, Scorza R, Della Rossa A, Bencivelli W. Assessment of disease severity and prognosis. *Clin Exp Rheumatol* 2003;21(3 Suppl 29):S42-6.
12. Tomarken AJ, Waller NG. Potential problems with "well fitting" models. *J Abnorm Psychol* 2003;112:578-98.
13. Hu L, Bentler P,M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model* 1999;6:1-55.
14. Mayes MD, Lacey JV,Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum* 2003;48:2246-55.

15. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd ed. New York: McGraw-Hill; 1994.
16. Blalock SJ, DeVellis RF, Brown GK, Wallston KA. Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations. *Arthritis Rheum* 1989;32:991-7.
17. Callahan LF, Kaplan MR, Pincus T. The Beck Depression Inventory, Center for Epidemiological Studies Depression Scale (CES-D), and General Well-Being Schedule depression subscale in rheumatoid arthritis. Criterion contamination of responses. *Arthritis Care Res* 1991;4:3-11.
18. Khanna D. Health-related quality of life: A primer with a focus on scleroderma. *Scleroderma Care and Research* 2006;32:3-13.

ROLE OF THE SPONSORS

The authors of this article had access to all study data; are responsible for all contents of the article, and had authority over manuscript preparation and the decision to submit the manuscript for publication. The sponsors of the Canadian Scleroderma Research Group did not have any role in the development of this manuscript or the decision to submit it for publication.

Table 1. Factor Loadings of Center for Epidemiological Studies Depression Scale Items in Models Tested

CES-D Item	Model 1:	Model 2:	Model 3a:	Model 3b:	Models 4a/4b:
	1-factor	2 Factors	3-factors (DA + S/V)	3-factors (DA + PA)	
1. I was bothered by things that usually don't bother me.	Depression = .70	DA + S/V + IP = .72	DA + S/V = .72	S/V = .77	S/V = .76
2. I did not feel like eating; my appetite was poor.	Depression = .55	DA + S/V + IP = .58	DA + S/V = .58	S/V = .60	S/V = .61
5. I had trouble keeping my mind on what I was doing.	Depression = .67	DA + S/V + IP = .69	DA + S/V = .69	S/V = .73	S/V = .73
7. I felt that everything I did was an effort .	Depression = .65	DA + S/V + IP = .68	DA + S/V = .68	S/V = .71	S/V = .71
11. My sleep was restless.	Depression = .45	DA + S/V + IP = .49	DA + S/V = .49	S/V = .50	S/V = .51
13. I talked less than usual.	Depression = .68	DA + S/V + IP = .70	DA + S/V = .70	S/V = .74	S/V = .74
20. I could not get going .	Depression = .69	DA + S/V + IP = .72	DA + S/V = .72	S/V = .76	S/V = .76
3. I felt that I could not shake off the blues ...	Depression = .83	DA + S/V + IP = .84	DA + S/V = .84	DA + PA = .84	DA = .85
6. I felt depressed .	Depression = .88	DA + S/V + IP = .89	DA + S/V = .89	DA + PA = .89	DA = .90
9. I thought my life had been a failure .	Depression = .78	DA + S/V + IP = .79	DA + S/V = .79	DA + PA = .80	DA = .80
10. I felt fearful .	Depression = .66	DA + S/V + IP = .68	DA + S/V = .68	DA + PA = .67	DA = .68

14. I felt <i>lonely</i> .	Depression = .77	DA + S/V + IP = .78	DA + S/V = .78	DA + PA = .78	DA = .79
17. I had <i>crying</i> spells.	Depression = .80	DA + S/V + IP = .81	DA + S/V = .81	DA + PA = .81	DA = .81
18. I felt <i>sad</i> .	Depression = .88	DA + S/V + IP = .89	DA + S/V = .90	DA + PA = .90	DA = .90
4. I felt that I was just as <i>good</i> as other people.	Depression = .37	PA = .63	PA = .63	DA + PA = .39	PA = .64
8. I felt <i>hopeful</i> about the future.	Depression = .56	PA = .79	PA = .79	DA + PA = .58	PA = .79
12. I was <i>happy</i> .	Depression = .68	PA = .91	PA = .91	DA + PA = .70	PA = .91
16. I <i>enjoyed</i> life.	Depression = .66	PA = .87	PA = .87	DA + PA = .68	PA = .86
15. People were <i>unfriendly</i> .	Depression = .63	DA + S/V + IP = .65	IP = .79	IP = .79	IP = .79
19. I felt that people <i>disliked</i> me.	Depression = .73	DA + S/V + IP = .74	IP = .93	IP = .93	IP = .93

Item allocation notations are based on Radloff's (1977) 4-factor model (DA = depressed affect, S/V = somatic/vegetative, PA = positive affect, IP = interpersonal). When two or more of Radloff's original factors are combined into a single factor, this is noted with an addition sign following Rhee et al. (5). For example, a single factor based on Radloff's DA and PA factors is shown as DA + PA. Factor loadings shown for the 4-factor models are from Model 4a, which were not substantively different than those for Model 4b.

Table 2. Fit Indices, Factor Correlations (Model 4a) and Second-Order Factor Loadings (Model 4b) for Confirmatory Factor Analysis Models

Model Fit Indices	χ^2	<i>df</i>	CFI	TLI	RMSEA	SRMR
Model 1a: 1 Factor	1079.4	58	.70	.86	.19	.14
Model 2: 2-Factor (DA + S/V + IP, PA)	243.8	67	.95	.98	.08	.06
Model 3a: 3-Factor (DA + S/V, PA, IP)	222.9	67	.96	.98	.07	.06
Model 3b: 3-Factor (DA + PA, S/V, IP)	973.3	57	.73	.87	.19	.13
Model 4a: 4-Factor (DA, S/V, IP, PA)	180.2	67	.97	.99	.06	.05
Model 4b: 4-Factor, 2 nd Order (DA, S/V, IP, PA)	180.5	67	.97	.99	.06	.06

Factor Correlations and Second-Order Factor Loadings	DA	S/V	PA	IP
Model 4a – Correlated 4-Factor:				
DA	-----			
S/V	0.89	-----		

PA	0.41	0.28	-----	
IP	0.77	0.69	0.29	-----
Model 4b - 2nd Order 4- Factor:				
Second-Order Factor Loadings	0.99	0.86	0.37	0.76

Figure 1

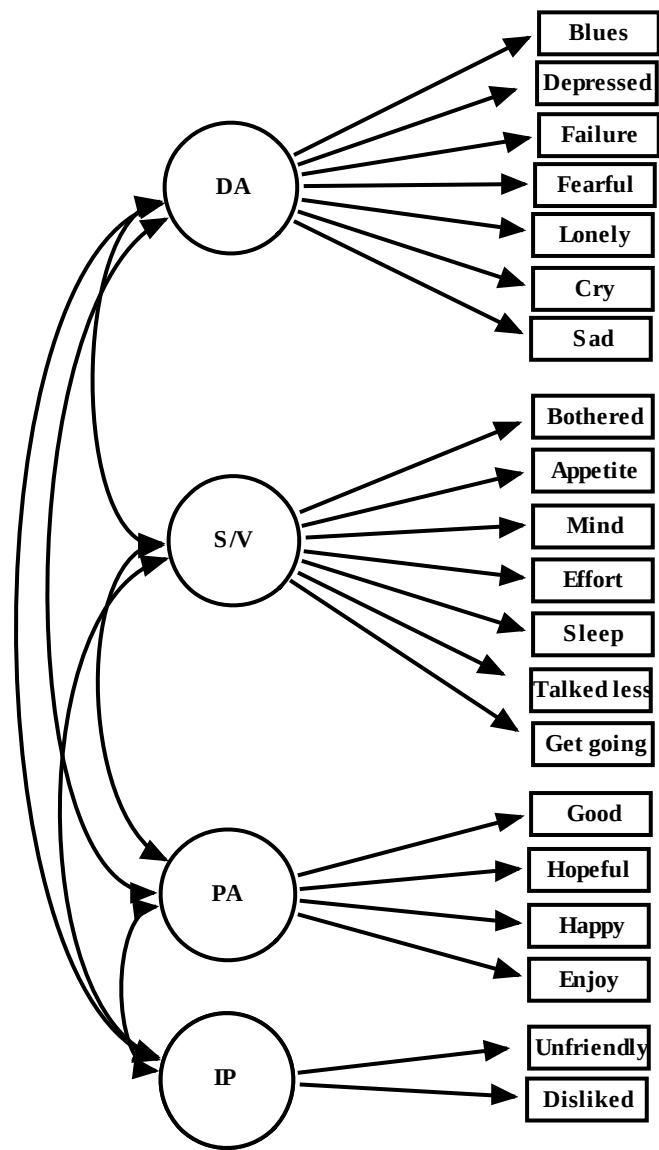


FIGURE LEGEND

Figure 1. Correlated 4-factor model with the Radloff (1977) item allocation. Item error variances are not shown. DA = Depressed Affect, S/V = Somatic/Vegetative, PA = Positive Affect, IP = Interpersonal.