



- ◆ Trabajo realizado por la Biblioteca Digital de la Universidad CEU-San Pablo
- ◆ Me comprometo a utilizar esta copia privada sin finalidad lucrativa, para fines de investigación y docencia, de acuerdo con el art. 37 de la M.T.R.L.P.I. (Modificación del Texto Refundido de la Ley de Propiedad Intelectual del 7 julio del 2006)

## Measurement of Depression: Comparison Between Self-Reports and Clinical Assessments of Depressed Outpatients

Aquilino Polaino<sup>1</sup> and Carmen Senra,<sup>2,3</sup>

Accepted: November 7, 1991

---

*Agreement among several depression scales was investigated as regards the relative influences of administration mode (self-rating or clinical rating) and scale content. The Beck Depression Inventory (BDI), the Self-Rating Depression Scale (SDS), the Hamilton Rating Scale for Depression (HRSD), and three corresponding scales with identical structure and content but the alternative administration mode were administered to 47 outpatients with diagnoses of DSM-III major depression disorders. Correlations between the total scores and the degrees of association between corresponding items of different scales were calculated. The results suggest that differences in content contribute more to inter-scale discrepancy than differences in administration mode. The implications for the evaluation of outpatients with major depression are discussed.*

---

**KEY WORDS:** Beck Depression Inventory (BDI); Self-Rating Depression Scale (SDS); Hamilton Rating Scale for Depression (HRSD); interscale agreement.

### INTRODUCTION

Depression is one of the major challenges to world health (Sartorius, 1979). Because of the high risk associated with undetected depression and the need for exact diagnosis and objective quantification of depressive

<sup>1</sup>Departamento de Personalidad, Evaluación y Psicología Clínica, Universidad Complutense de Madrid, Madrid, Spain.

<sup>2</sup>Departamento de Psicología Clínica y Psicobiología, Facultad de Psicología, Universidad de Santiago de Compostela, Coruña, Spain.

<sup>3</sup>To whom correspondence should be addressed.

symptoms, one of the key areas of research on this disorder concerns the instruments by which depression is evaluated (Mayer, 1977; Polaino-Loriente, 1985). Over the years, however, the proliferation of instruments for the evaluation of depressive symptomatology, self-report inventories and interviewer rating scales in particular, has led to clinicians and researchers being faced with the problem of deciding which of the numerous available instruments is best suited for a given task (Lambert, Hatch, Kingston, & Edwards, 1986).

Three of the most widely used instruments for measuring the intensity of depressive symptoms are the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) and the Self-Rating Depression Scale (SDS; Zung, 1965). Correlation between the total BDI and the total SDS scores of psychiatric patients is high and significant (Davies, Burrows, & Poynton, 1975; Seitz, 1970; Schnurr, Hoaken, & Jarret, 1976); correlation between HRSD and SDS scores has been reported as either nonsignificant (White, White, & Razani, 1984) or significant but low (Biggs, Wylie, & Ziegler, 1978; Carroll, Fielding, & Blashki, 1973; Schnurr, *et al.*, 1976); and correlation between the BDI and the HRSD has been variously found to be nonsignificant (Schnurr *et al.*, 1976), significant but low (Bailey & Copen, 1976; Post *et al.*, 1985), moderate (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981; Fitzgibbon, Cella, & Sweeney, 1988; Steer, Beck, Riskind, & Brown, 1987), or high (Davies *et al.*, 1975; Williams, Barlow, & Agras, 1972).

In general, the above results exhibit (a) wide differences among the coefficients of correlation reported, presumably due to the differences among the samples studied and the measures of correlation used; (b) a better overall correlation between the self-rating scales than between these and the clinical rating scale; and (c) a better overall correlation between the BDI and the HRSD than between the SDS and the HRSD.

Some of those who found poor a correlation between self-rating and clinical rating scales (Arfwidson *et al.*, 1974; Carroll *et al.*, 1973; Prusoff, Klerman, & Paykel, 1972) opined that, although neither method can be declared superior to the other, use of the self-rating method should be restricted to case finding or as a supplementary test. There is evidence, however, that the lack of correlation derives not only from the difference in administration mode, but also from differences with regard to the item content, the time frame of the rating, and response structure (Wetzler, Kahn, Strauman, & Dubro, 1989). Moran and Lambert (1983), for example, found that the BDI, SDS, and HRSD measure different aspects of depression as defined by the DSM-III, and Steer *et al.*'s (1987) analysis of the separate and combined factor structures of the BDI and HRSD

suggested not only that scores on the two scales differ because of the difference between the raters' viewpoints, but also that the scales themselves concentrate on different clusters of symptoms. However, the need to study the discrepancy between self-raters and clinicians in the absence of differences in response structure (e.g., intensity vs. frequency) and item content (Lambert et al., 1986; Prusoff *et al.*, 1972) has not yet been met.

The purpose of the present study was to examine the influence of administration mode and content differences on the discrepancies among BDI, SDS and HRSD scores for overall intensity of depression and specific symptoms. Bearing in mind that depression scales can detect and evaluate depressive symptoms (dysphoria) but not diagnose major depression disorder (Hagan and Schauer, 1985; Hesselbrock, Tennen, Hesselbrock, Workman, & Meyer, 1985; Zich, Attkisson, & Greenfield, 1990), outpatients with DSM-III diagnoses of major depression were administered the BDI, SDS, and HRSD scales and also corresponding forms with identical structure and content but the alternative administration mode.

## METHOD

### Subjects

The study was carried out in the Neuropsychiatric Outpatient Service of the La Robleda Psychiatric Hospital, Santiago de Compostela, Spain. The sample comprised 29 patients with DSM-III single-episode major depression disorders and 18 patients with recurrent-episode major depression disorders; 32 (68%) were women and 15 (32%) men. Their mean age was 34.4 years ( $SD = 8.23$  years). For most patients, this was their first or second depressive episode, and the episodes had begun a mean 4.1 months before presentation in our clinic. No patient had a diagnosis of DSM-III bipolar or melancholia disorders, none had a history of addiction other than to tobacco, all were drug-free for at least 2 weeks before being interviewed, and all were willing and able to complete the inventories.

The male/female ratio among first-episode patients was similar to that among recurrent-episode patients [ $\chi^2(1) = .02$ , n.s.]. Student's *t* comparison of total scores for the standard forms of the BDI, SDS, and HRSD (see Table I for means and SD) revealed no significant differences between these two groups:  $t(45) = .36$  (n.s.) for the BDI  $t(45) = .43$  (n.s.) for the SDS, and  $t(45) = .1$  (n.s.) for the HRSD. The two groups were accordingly pooled for subsequent analyses.

Table 1. Mean BDI, SDS, and HRSD of Single-Episode and Recurrent-Episode Groups

	BDI		SDS		HRSD	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Single episode ( <i>n</i> = 29)	25.31	3.76	51.00	2.58	20.58	2.35
Recurrent episode ( <i>n</i> = 18)	25.72	3.99	51.33	2.74	21.22	1.87

### Instruments

*Semistructured Diagnostic Interview (SDI).* Patients were selected by means of an ad hoc semistructured interview (Senra, 1990) that both fulfilled DSM-III guidelines for diagnosis of major depression disorders and included items of the interview forms of the scales being investigated. The SDI consisted of five sections recording (1) personal data, (2) essential signs and symptoms, (3) concomitant signs and symptoms, (4) disqualifying symptoms, and (5) the profile of the episode.

*The Beck Depression Inventory (BDI).* The BDI quantifies the severity of 21 depressive symptoms. Though originally designed for interviewer-assisted completion (1961 version), it is now generally used in the self-report mode. Each of the 21 items is composed of four alternative statements rating the severity of a symptom from 0 to 3. The reliability and validity of the BDI for measuring the intensity of depressive symptoms have been confirmed by Bech (1981), Beck and Beamesderfer (1974), Beck, Steer, and Garbin (1988), Mayer (1977), Rehm (1976), and Steer, Beck, and Garrison (1986). However, Schwab, Bialow, Brown, and Holzer (1967) asserted that the BDI is biased towards pessimism, failure, and self-punitive wishes, whereas the somatic dimension contributes only 29% to the total score. The Spanish version of the BDI was developed and validated by Conde and Useros (1974) and Conde, Esteban, and Useros (1976).

In this study we used both the usual 21-item self-rating version of the inventory (B-SR) and an interview form of identical content and scoring structure (B-CI).

*The Zung Self-Rating Depression Scale (SDS).* The SDS, designed "to quantify depression as a disorder" (Zung, 1965), comprises 20 items scored 1-4 according to symptom frequency. The SDS gives considerable weight to symptoms involving physiological dysfunction (Rehm, 1976); somatic and behavioral symptoms together make up 50% of the items, bringing its content more into line with the Hamilton scale than is the BDI (Carroll *et al.*, 1973). Studies by Biggs *et al.* (1978), Hedlund and Vieweg (1979a), and Zung (1986) testify to the reliability and validity of the SDS, which is widely

accepted for assessing the general severity of depression, but Mayer (1977), Moran and Lambert (1983), and Rehm (1976) insist on the need for further research on its internal consistency and discriminant validity. The Spanish version has been developed and validated by Conde and co-workers (Conde, Escriba, & Izquierdo, 1970a, b; Conde & Esteban, 1973; Conde & Sánchez, 1969).

In this study we used both the usual self-rating form of the scale (S-SR) and an interview form of identical content and scoring structure (S-CI; Conde and Franch, 1984).

*The Hamilton Rating Scale for Depression (HRSD).* The HRSD is designed to quantify the results of clinical interviews of patients with diagnoses of depressive affective disorders (Hamilton, 1960, 1967). The original version contained 21 items, but only 17 were considered necessary for evaluation of the overall severity of depression, the omitted items being diurnal variation, derealization, paranoid symptoms, and obsessional symptoms (Hamilton, 1960). The items have either 3- or 5-point ratings for severity. Carroll *et al.* (1973) pointed out that somatic and behavioral symptoms account for 50-80% of the total score. The interrater reliability of the HRSD and its validity as a measure of the severity of symptoms of depression have been confirmed by Hamilton (1986), Hedlund and Vieweg (1979b), Mayer (1977), Moran and Lambert (1983), and Rehm and O'Hara (1985). Its administration has been standardized by the introduction of the Structured Interview Guide for the HRSD (SIGH-D; Williams, 1988).

In this study we used both the usual 17-item clinical interview form of the scale (H-CI) and a self-rating form of identical content and scoring structure (H-SR).

### Procedure and Data Analysis

Before patients were selected for the study, interrater reliability was evaluated by having the two raters jointly interview 20 patients not included in the study and (independently) rate them on the three interview forms. The interrater product-moment correlations were .80 for the SDS, .89 for the HRSD, and .93 for the BDI.

The patients were tested in the course of the routine evaluation procedure of the Outpatient Service. They were told that the research was aimed at understanding their symptoms and that it was important for them to report their symptoms as accurately as possible. Patient selection began with the SDI. The 51 patients fulfilling initial criteria for inclusion in the study were then interviewed on the HRSD. For patients scoring 14 or more on the full HRSD (no patient was in fact excluded by this threshold), interviewing con-

tinued with the interview forms of the BDI and the SDS. The patients then started self-reporting in a quiet room under the supervision of a psychologist who was experienced in the use of these scales but had no knowledge of the patients' interview ratings. The first self-report scale administered was the BDI; only patients scoring 18 or more on this scale (all except two) went on to complete the SDS and the self-report form of the HRSD. Both the self-report and the interview forms of the three scales were administered at the time of admission, and all referred to the patient's state in "the past week, including today." All scales were administered in the morning to minimize differences associated with circadian variations. After scale administration, an experienced psychiatrist ignorant of the scale ratings had a routine diagnostic interview with the patients and prescribed medication. The final requisite for a patient's inclusion in the study was that the psychiatrist should diagnose DSM-III major depression; only two patients were excluded by this criterion. The number of patients who completed all the scales and were definitively included in the study was therefore 47.

Pearson product-moment correlations among the total scores of the six scales used were calculated from the data for the 47 subjects. The significance and degree of association among the items of the CI and SR forms of a given scale, and among the items of similar content in different scales, were estimated using chi-square tests and contingency coefficients. The Bonferroni-Holm procedure was used to correct significance levels.

## RESULTS

### Agreement Among Total Scores

Table II lists the Pearson correlations among the total scores of the six scale forms used. At the  $p < .001$  significance level, there is a better correlation between the standard (SR) forms of the Beck and Zung scales than between these and the standard CI form of the Hamilton scale ( $r = .60$  for B-SR/S-SR,  $r = .50$  for B-SR/H-CI,  $r = .47$  for S-SR/H-CI). The best correlations are between the two forms of the same scale:  $r = .71$  for B-SR/B-CI,  $r = .65$  for S-SR/S-CI, and  $r = .62$  for H-SR/H-CI ( $p < .001$  in all three cases).

### Agreement Among Symptoms

Table III lists the groups of items considered as comparable on the three scales. Table IV shows the contingency coefficients relating corre-

Table II. Pearson Correlations Among the SR/CI Forms of the Beck, Zung, and Hamilton Scales ( $N = 47$ )

	B-SR	B-CI	S-SR	S-CI	H-SR	H-CI
B-SR	—					
B-CI	.71*	—				
S-SR	.60*	.52*	—			
S-CI	.50*	.61*	.65*	—		
H-SR	.54*	.44*	.51*	.47*	—	
H-CI	.50*	.56*	.47*	.53*	.62*	—

\*  $p < .001$ .

Table III. Correspondences Among BDI, SDS, and HRSD Items

Symptom	Beck	Zung	Hamilton
Sadness	1	1	Depressed mood
Crying	10	3	
Pessimism	2	14	
Uselessness	3	17	1
Guilt, punishment, self-accusation	5,6,8		2
Insomnia	16	4	4,5,6
Loss of appetite	18	5,8	12
Loss of libido	21	6	14
Weight loss	19	7	16
Suicidal ideas	9	19	3
Indecisiveness	13	16	
Fatigability	17	10	13
Hypochondriasis	20		15
Irritability	11	15	10
Somatic anxiety		9	11
Agitation		13	9
Retardation		12	8
Work and interest	4,15	20,18	7

spending items of the two forms of each scale (columns 2-4), the pairwise associations between corresponding items of the three self-report scales (columns 5-7) and the associations between corresponding items of the HRSD-CI and the standard self-report scales (columns 8 and 9).

The general pattern of these results repeats that obtained for the total scores. For a given symptom, the association between the two forms of a given scale (columns 2-4) is generally greater than the association between different scales (regardless of the forms used), the chief exceptions being loss of libido and indecisiveness. The smallest contingency coefficients are almost without exception those for comparisons between the H-CI and the



Table IV. Contingency Coefficients Between Corresponding Items of Different Scale Forms

Symptoms	B-SR/B-CI	S-SR/S-CI	H-SR/H-CI	B-SR/S-SR	B-SR/H-SR	S-SR/H-SR	H-CI/B-SR	H-CI/S-SR
Sadness	.63***	.59***		.58***				
Crying	.65***	.62**		.61**				
Pessimism	.61***	.55**	.59***	.57*	.55	.53	.50	.49
Uselessness	.57*	.62**		.55*				
Guilt	.48*							
Punishment	.52*		.49		.47		.47	
Self-accusation	.54**							
Insomnia	.70***	.70***	.76***	.64***	.65***	.67***	.60***	.60***
Loss of appetite	.63***	.63***	.62***	.60***	.60***	.57**	.54**	.53*
Loss of libido	.53**	.53*	.53**	.59***	.58***	.59**	.49*	.48*
Weight loss	.69***	.68***	.68***	.65***	.63***	.63***	.60***	.60***
Suicide	.70***	.67***	.66***	.65***	.64***	.62***	.62***	.60***
Indecisiveness	.54*	.52		.60**				
Fatigability	.63***	.62***	.61***	.60**	.56***	.58**	.51*	.55*
Hypochondria	.47		.51		.41		.40	
Irritability	.57**	.52	.53	.50	.49	.48	.42	.43
Somatic anxiety		.53	.55			.50		.43
Agitation		.41	.38			.35		.26
Retardation		.44	.43			.40		.29
Work and interest	.60***	.56*	.57**	.55*	.59*	.51	.56*	.40

\*  $p \leq .05$ , as corrected by the Bonferroni-Holm procedure.\*\*  $p \leq .01$ , as corrected by the Bonferroni-Holm procedure.\*\*\*  $p \leq .001$ , as corrected by the Bonferroni-Holm procedure.

standard self-report scales (columns 8 and 9). However, the trends among the contingency coefficients of the various items are very similar for all three kinds of comparison, with the greatest association for insomnia, suicide, and weight loss and the least for feelings of guilt, hypochondriasis, anxiety, agitation, and psychomotor retardation. In fact, the omission of agitation and retardation from the Zung and Hamilton scales increases the SR/CI agreement between the total scores of these scales from .62 to .72 for the HRSD and from .65 to .73 for the SDS.

## DISCUSSION

In keeping with the general pattern observed in the literature, our total score results show a better correlation between the BDI and the SDS than between either of these scales and the HRSD. However, the finding that all three single-scale SR/CI pairs exhibited a better correlation than any mixed-scale pair appears to be good evidence that the differences between results obtained with the standard forms of the scales are due more to differences in the structure and content of the scales than to discrepancies in administration mode; although the order of application of the two modes may have increased the apparent intermethod agreement due to the clinician having possibly altered the patient's preexisting attitude to certain symptoms, this influence would also have affected the other comparisons involving such symptoms.

We also found better correlation between the BDI and the HRSD than between the latter and the SDS, in spite of the content of the SDS being closer than that of the BDI to the content of the HRSD. This may have been because the SDS measures the frequency of symptoms, while the other two scales measure symptom intensity; in other words, the interscale agreement may have been enhanced not only by the similarity of content but also by the specific way in which content is elicited. This effect (an inevitable side-effect of ensuring equal content) would presumably be even more marked between the two forms of the same scale, which have items phrased in almost exactly the same way.

Our results for individual symptoms largely reflect those for the total scores: the association between the two forms of the same scale was greater than that between different scales (the exceptional behaviour of loss of libido and indecisiveness is probably due to uncertainty about the patient's basal behaviour in these respects). For most items the agreement between SR and CI forms was better for the Beck scale than for either of the others, possibly because the BDI, which emphasizes the cognitive-affective dimen-

sion of depression, fits in better with the patient's own subjective awareness of his or her condition.

Our results for individual symptoms confirm the conclusions of Paykel and Norton (1986) concerning the difficulty patients have in quantifying symptoms such as hypochondriasis, feelings of guilt, agitation, and retardation. It should be borne in mind, moreover, that the DSM-III and most clinical interview protocols stipulate that the patient's verbal report is an insufficient basis for the quantification of agitation and retardation, which should take into account the patient's behavior throughout the interview. In view of this stipulation, and since in this study the omission of these two symptoms considerably increased overall SR/CI agreement for the Zung and Hamilton scales, we believe that the omission of these symptoms from self-rating scales should be considered.

The results of this study corroborate the utility of self-report scales for quantitating the severity of symptoms of depression in outpatients with major depression but no psychotic or melancholia disorders. Note, however, that our findings apply strictly to the Spanish versions of these scales. It may also be pointed out that differences in content are extremely relevant to the efficacy of different scales in monitoring patients' progress under treatment, since the main effects of different therapies involve different aspects of depression. Future research should tackle the influence of scale administration mode and content on the evaluation of patients with other diagnoses and on the monitoring of changes accompanying therapy.

## REFERENCES

- Arfwidsson, L., D'Elia, G., Laurell, B., Ottosson, J., Perris, C., & Persson, G. (1974). Can self-rating replace doctor's rating in evaluating anti-depressive treatment? *Acta Psychiatrica Scandinavica*, *50*, 16-22.
- Bailey, J., & Coppen, A. (1976). A comparison between the Hamilton Rating Scale and Beck Inventory in the measurement of depression. *British Journal of Psychiatry*, *128*, 486-489.
- Beck, P. (1981). Rating scales for affective disorders: Their validity and consistency. *Acta Psychiatrica Scandinavica*, *64* (Suppl. 295), 17-31.
- Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: The depression inventory. In P. Pichot (Ed.), *Psychological measurements in psychopharmacology. Modern problems of pharmacopsychiatry* (pp. 151-169). Basel, Switzerland: Karger.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561-571.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Psychological Review*, *8*, 77-100.
- Biggs, J. T., Wylie, L. T., & Ziegler, V. E. (1978). Validity of the Zung Self-Rating Depression Scale. *British Journal of Psychiatry*, *132*, 381-385.
- Carroll, B. J., Fielding, J. M., & Blashki, T. G. (1973). Depression rating scales. *Archives of General Psychiatry*, *28*, 361-366.

- Carroll, B. J., Feinberg, M., Smouse, P. E., Rawson, S. G., & Greden, J. F. (1981). The Carroll Rating Scale for Depression. I. Development reliability and validity. *British Journal of Psychiatry*, *138*, 194-200.
- Conde, V., & Esteban, T. (1973). Revisión crítica de dos adaptaciones castellanas de la "Self-Rating Depression Scale (SDS)" de Zung. *Archivos de Neurobiología*, *34*, 375-392.
- Conde, V., & Franch, J. I. (1984). *Escalas de evaluación comportamental para la cuantificación de la sintomatología psicopatológica en los trastornos angustiosos y depresivos*. Madrid: Upjohn.
- Conde, V., & Sánchez, J. (1969). La escala autoaplicada para la depresión de Zung. *Archivos de Neurobiología*, *33*, 535-558.
- Conde, V., & Useros, E. (1974). El inventario para la medida de la depresión de Beck. *Revista de Psiquiatría y Psicología Médica de Europa y América Latina*, *12*, 153-167.
- Conde, V., Escriba, J., & Izquierdo, J. (1970a). Estudio crítico de la escala autoaplicada para la medida cuantitativa de la depresión de Zung. *Revista Clínica Española*, *117*, 71-78.
- Conde, V., Escriba, J., & Izquierdo, J. (1970b). Evaluación estadística y adaptación castellana de la escala autoaplicada para la depresión de Zung. *Archivos de Neurobiología*, *33*, 281-303.
- Conde, V., Esteban, T., & Useros, E. (1976). Revisión crítica de dos adaptaciones castellanas del cuestionario de Beck. *Revista de Psicología General y Aplicada*, *31*, 469-497.
- Davies, B., Burrows, G., & Poynton, C. (1975). A comparative study of depression rating scales. *Australian and New Zealand Journal of Psychiatry*, *9*, 21-24.
- Fitzgibbon, M. L., Cella, D. F., & Sweeney, J. A. (1988). Redundancy in measures of depression. *Journal of Clinical Psychology*, *44*, 372-374.
- Hagan, L. D., & Schauer, A. H. (1985). Assessment of depression in alcoholics: Comment on Hesselbrock et al. *Journal of Consulting and Clinical Psychology*, *53*, 64-66.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56-62.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, *6*, 278-296.
- Hamilton, M. (1986). The Hamilton Rating Scale for Depression. In N. Sartorius & T. Ban (Eds.), *Assessment of depression* (pp. 143-152). Berlin: Springer-Verlag.
- Hedlund, J. L., & Vieweg, B. W. (1979a). The Zung Self-Rating Depression Scale: A comprehensive review. *Journal of Operational Psychiatry*, *10*, 51-64.
- Hedlund, J. L., & Vieweg, B. W. (1979b). The Hamilton Rating Scale for Depression: A comprehensive review. *Journal of Operational Psychiatry*, *10*, 149-165.
- Hesselbrock, M., Tennen, H., Hesselbrock, V., Workman, K., & Meyer, R. (1985). Assessment of depression in alcoholics: Further considerations-Reply to Hagan and Schauer. *Journal of Consulting and Clinical Psychology*, *53*, 67-69.
- Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton rating scales as measures of treatment outcome: A meta-analytic comparison. *Journal of Consulting and Clinical Psychology*, *54*, 54-59.
- Mayer, J. M. (1977). Assessment of depression. In P. McReynold (Ed.), *Advances in psychological assessment*, Vol. 4 (pp. 358-425). Washington DC: Jossey Bass.
- Moran, P. W., & Lambert, M. J. (1983). A review of current assessment tools for monitoring changes in depression. In M. J. Lambert, E. R. Christensen, S. S. DeJulio (Eds.), *The measurement of psychotherapy outcome in research and evaluation* (pp. 263-303). New York: Wiley.
- Paykel, E. S., & Norton, K. R. W. (1986). Self-report and clinical interview in the assessment of depression. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 356-366). Berlin: Springer-Verlag.
- Polaino-Lorente, A. (1985). *La depresión*. Barcelona: Martínez Roca.
- Post, R. D., Alford, C. E., Baker, N. J., Franks, R. D., House, R. M., Jackson, A. M., & Petersen, J. L. (1985). Comparison of self-reports and clinicians' ratings of unipolar major depression. *Psychological Reports*, *57*, 479-483.

- Prusoff, B. A., Klerman, G. L., & Paykel, E. S. (1972). Concordance between clinical assessment and patients' self-report in depression. *Archives of General Psychiatry, 26*, 546-552.
- Rehm, L. P. (1976). Assessment of depression. In M. Hersen & A. Bellack (Eds.), *Behavioral assessment: A practical handbook* (pp. 233-259). New York: Pergamon Press.
- Rehm, L. P., & O'Hara, M. W. (1985). Item characteristics of the Hamilton Rating Scale for Depression. *Journal of Psychiatry Research, 19*, 31-41.
- Sartorius, N. (1979). Research on affective psychoses with the framework of the WHO programme. In M. Schon & E. Strömngren (Eds.), *Origin, prevention and treatment of affective disorders* (p. 207). London: Academic Press.
- Schnurr, R., Hoaken, P. C. S., & Jarret, F. J. (1976). Comparison of depression inventories in a clinical population. *Canadian Psychiatric Association Journal, 21*, 473-476.
- Schwab, J. J., Bialow, M., Brown, J. M., & Holzer, C. E. (1967). Diagnosing depression in medical patients. *Annals of Internal Medicine, 67*, 695-707.
- Seitz, F. C. (1970). Five psychological measures of neurotic depression: A correlation study. *Journal of Clinical Psychology, 26*, 504-505.
- Senra, C. (1990). Semistructured diagnostic interview for DSMIII major depression (available from Dr. C. Senra, Dpto. Psicología Clínica, 15705-Santiago de Compostela, La Coruña, Spain).
- Steer, R. A., Beck, A. T., & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 123-142). Berlin: Springer-Verlag.
- Steer, R. A., Beck, A. T., Riskind, J. H., & Brown, G. (1987). Relationships between the Beck Depression Inventory and the Hamilton Psychiatric Rating Scale for Depression in depressed outpatients. *Journal of Psychopathology and Behavioral Assessment, 9*, 327-339.
- Wetzler, S., Kahn, R., Strauman, T., & Dubro, A. (1989). Diagnosis of major depression by self-report. *Journal of Personality Assessment, 53*, 22-30.
- White, J., White, K., & Razani, J. (1984). Effects of endogeneity and severity on consistency of standard depression rating scales. *Journal of Clinical Psychiatry, 45*, 260-261.
- Williams, J. B. W. (1988). A Structured Interview Guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry, 45*, 742-747.
- Williams, J. G., Barlow, D. H., & Agras, W. S. (1972). Behavioural measurement of severe depression. *Archives of General Psychiatry, 27*, 330-333.
- Zich, J. M., Attkisson, C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: The CES-D and the BDI. *International Journal of Psychiatry in Medicine, 20*, 259-277.
- Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12*, 63-70.
- Zung, W. W. K. (1986). Zung Self-rating Depression Scale and Depression Status Inventory. In N. Sartorius & T.A. Ban (Eds.), *Assessment of depression* (pp. 221-231). Berlin: Springer-Verlag.