

- Hankin, B. L., Fraley, R. C., Lahey, B. B., & Waldman, I. D. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology, 114*, 96–110.
- Joiner, T. E., Jr., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment, 17*, 267–277.
- Kendall, P. C., & Flannery-Schroeder, E. C. (1995). Rigor, but not rigor mortis, in depression research. *Journal of Personality and Social Psychology, 68*, 892–894.
- Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding the use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11*, 289–299.
- Kumar, G., Steer, R. A., Teitleman, K. B., & Villacis, L. (2002). Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment, 9*, 164–170.
- Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology, 111*, 290–301.
- Santor, D. A., & Coyne, J. C. (2001). Evaluation of the continuity of symptomatology between depressed and nondepressed individuals. *Journal of Abnormal Psychology, 110*, 216–225.
- Slade, T., & Andrews, G. (2005). Latent structure of depression in a community sample: A taxometric analysis. *Psychological Medicine, 35*, 489–497.
- Tennen, H., Hall, J. A., & Affleck, G. (1995). Depression research methodologies in the Journal of Personality and Social Psychology: A review and critique. *Journal of Personality and Social Psychology, 68*, 870–884.
- Waller, N. G., Compas, B. E., Hollon, S. D., & Beckjord, E. (2005). Measurement of depressive symptoms in women with breast cancer and women with clinical depression: A differential item functioning analysis. *Journal of Clinical Psychology in Medical Settings, 12*, 127–141.

## The Future of Depression Measurement Research

Leonard J. Simms

*Department of Psychology*

*University at Buffalo, The State University of New York*

As described by Santor, Gregus, and Welch (this issue) in their focus article, a number of interesting patterns have emerged over the last 80 years with respect to the measurement of depression severity. Despite the fact that Santor et al. identified nearly 300 measures of depression severity in their review of the literature, relatively few measures are used today in both basic science and treatment

---

Correspondence should be addressed to Leonard J. Simms, Department of Psychology, Park Hall 218, University at Buffalo, The State University of New York, Buffalo, NY 14260. E-mail: ljsimms@buffalo.edu

studies of depression. Moreover, these few prominent measures—such as the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), its second edition (BDI-II; Beck, Steer, & Brown, 1996), the Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960), and the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977)—generally are 30 years old or older (with the exception of the BDI-II) and continue to be used with great regularity despite differences in their symptom coverage and in their convergence with existing diagnostic systems, such as the criteria for major depressive disorder outlined by the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000)*.

In this comment, I focus on two general issues raised recently in the depression measurement literature and for which Santor et al.'s article provides the necessary context. In particular, the depression measurement literature has been limited in (a) its adherence to a single-dimension view of depressive symptomatology, and (b) its reliance on antiquated test development methods that have resulted in relatively inefficient measures with problematic psychometric features. I briefly review each of these points and provide recommendations for future depression measurement research.

### DIMENSIONALITY OF DEPRESSION

First, Santor et al.'s (this issue) data suggest that the predominant method for assessing depression today is to collect self-reported ratings of general depressive severity, neglecting the inherent multidimensionality of depression identified in many studies. For example, factor analyses of the BDI generally have revealed either two or three factors, but the factor structure has varied widely across studies. A variety of two- and three-factor models have been reported in the literature in which various combinations of cognitive, affective, and somatic item content have emerged as factors (e.g., Campbell, Burgess, & Finch, 1984; Steer, Ball, Ranieri, & Beck, 1999). Likewise, some factor analytic studies of the CES-D have suggested that four dimensions—depressed affect, positive affect, somatic disturbance, and interpersonal difficulties—underlie responses to that measure (e.g., Joseph & Lewis, 1995; Radloff, 1977), again providing evidence for the multidimensionality of depression. However, similar to what has been found in the BDI literature, other researchers have identified alternative structures underlying the CES-D (e.g., Foley, Reed, Mutran, & DeVellis, 2002; Helmes & Nielsen, 1998).

Thus, although factor analytic studies have identified a variety of candidate dimensions of depressive symptomatology, no consensus has emerged as to the optimal nature and number of dimensions that are needed to account for the variability seen in clinical presentations of depression. However, one clear conclusion to be drawn from these studies is that depression is multidimensional in nature; as such, measures of depression ought to reflect this multidimensionality through the use of

empirically based symptom subscales. Unfortunately, because they were designed to reflect a single underlying dimension of depression severity, the prominent measures of depression in use today likely do not contain enough breadth or depth of item content to yield replicable and meaningful factor analytic results. Thus, future work is needed to develop new measures of depression designed from the start to explore and reflect the multidimensionality of depression.

Likewise, as pointed out by Santor et al.'s (this issue) data, prominent depression measures continue to include a fair amount of specifically anxious content as well as symptoms that are related to both anxiety and depression (e.g., irritability, sleep disturbance, concentration difficulties). Unfortunately, including such content on purported measures of depression confounds the conclusions that can be drawn from data collected using these measures. Studies have revealed significant overlap among self-report measures of anxiety and depression (e.g., Watson et al., 1995), and such overlap also exists in the diagnostic criteria, resulting in substantial and, some would suggest, artificially high comorbidity rates for anxiety and depression (e.g., Kessler et al., 1996).

Recent models—such as the tripartite model of depression and anxiety (Clark & Watson, 1991) and extensions of it (Mineka, Watson, & Clark, 1998)—have suggested that many of the symptoms included on traditional depression severity measures are relatively nonspecific with respect to anxiety and depression. These nonspecific symptoms include both anxious and depressed mood, as well as other symptoms that are prevalent in both types of disorder, such as irritability, sleep difficulties, restlessness, and poor concentration. In addition to these nonspecific symptoms, the tripartite model suggests that anxiety and depression are differentiated by relatively specific dimensions of somatic arousal (e.g., shortness of breath, increased heart rate, dizziness, dry mouth) and anhedonia (e.g., absence of positive affect, loss of interest in usually pleasurable activities), respectively. This model has generated a fair amount of supporting evidence and may represent a reasonable foundation on which to build future measures of depressive symptomatology that differentiate the specific and nonspecific elements of the depression construct. Watson et al. (1995) introduced a measure—the Mood and Anxiety Symptom Questionnaire (MASQ)—that was designed specifically to tap the general and specific dimensions predicted by the tripartite model. However, as pointed out by Santor et al. (this issue), modern and theoretically informed measures of depression such as the MASQ have not yet gathered enough interest to supplant the historically prominent depression measures still in use today.

## USE OF MODERN MEASUREMENT THEORY

Although the prominent measures identified in Santor et al.'s (this issue) review—such as the BDI, HAM-D, and CES-D—largely have served the field well, they were developed without the benefit of modern psychometric models,

such as item response theory (IRT; Hambleton & Swaminathan, 1985; Lord, 1980). Recent studies using such modern methods have revealed a number of psychometric liabilities associated with these prominent depression measures. The use of IRT—a term that applies to a group of psychometric models that characterize psychological test items by one or more item parameters—as a tool to aid psychopathology and personality researchers has become increasingly popular over the past decade. IRT methods offer several advantages over classical test theory methods (Embretson, 1996), on which the development of most depression measures has been based to date. Relevant to the measurement of depression, the most notable advantages of IRT are (a) the ability to describe the points along a scale continuum where measurement is most and least precise, and (b) the opportunity to detect item bias, or differential item functioning (DIF; Holland & Wainer, 1993) across samples differing in important variables such as gender, ethnicity, or patient status.

For example, Santor, Ramsay, and Zurhoff (1994) examined gender item bias on the BDI using DIF techniques and identified small but significant amounts of bias in Items 6 (sense of punishment), 10 (crying), and 14 (distortion of body image). Likewise, Santor and Coyne (2001) identified several HAM-D symptoms that displayed DIF between nondepressed and clinically depressed adults experiencing similar levels of overall depressive severity. In particular, items tapping depressed mood, anhedonia, and suicidality were more likely to be expressed in depressed than in nondepressed individuals, whereas hypochondriasis and middle insomnia items were more likely to be expressed in nondepressed individuals at similar levels of depression severity. Similar studies conducted with the BDI (Bedi, Maraun, & Chrisjohn, 2001; Hammond, 1995) also have revealed multiple items that perform differentially across groups varying in patient status but with little consistency across studies. Together, these DIF studies suggest that the appropriateness of depression measures across samples varying in variables such as gender and patient status cannot be assumed; rather, it needs to be studied empirically.

The implications of this small sampling of studies are that (a) the psychometric characteristics of prominent measures of depression severity need to be studied more carefully using modern psychometric methods, and (b) new and/or modified measures of depressive symptomatology are needed that are built from the ground up using state-of-the-art psychometric procedures.

## CONCLUDING COMMENTS

Continued research examining the multidimensional nature of depression is needed to advance our understanding of depression in general and to inform the development of the next generation of depression measures. Single-dimension measures of depression—which predominate in current basic and applied de-

pression research—are outdated and should be replaced with measures informed by recent and future empirical work highlighting the dimensions underlying the phenotypic expression of depressive symptomatology. Moreover, modern psychometric principles should be used in these efforts to ensure that the resultant measures efficiently, precisely, and fairly measure the dimensions of depressive symptomatology across different populations of interest. One such measurement project—development of the Iowa Depression and Anxiety Scales (IDAS; Watson et al., 2006)—currently is under way and has resulted in a provisional measure including both general and specific scales of anxious and depressive symptomatology. However, given the challenges described by Santor et al. (this issue), convincing researchers and clinicians to adopt a new measure like the IDAS, or others like it, will be a difficult task given the inertia accompanying current use of prominent measures like the BDI, HAM-D, and CES-D in basic and applied settings.

## REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–569.
- Bedi, R. P., Maraun, M. D., & Chrisjohn, R. D. (2001). A multisample item response theory analysis of the Beck Depression Inventory-1A. *Canadian Journal of Behavioural Science, 33*, 176–185.
- Campbell, I., Burgess, P. M., & Finch, S. J. (1984). A factorial analysis of BDI scores. *Journal of Clinical Psychology, 40*, 992–996.
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*, 316–336.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*, 341–349.
- Foley, K. L., Reed, P. S., Mutran, E. J., & DeVellis, R. F. (2002). Measurement adequacy of the CES-D among a sample of older African-Americans. *Psychiatry Research, 109*, 61–69.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*, 56–61.
- Hammond, S. M. (1995). An IRT investigation of the validity of non-patient analogue research using the Beck Depression Inventory. *European Journal of Psychological Assessment, 11*, 14–20.
- Helmes, E., & Nielson, W. R. (1998). An examination of the internal structure of the Center for Epidemiological Studies-Depression Scale in two medical samples. *Personality and Individual Differences, 25*, 735–743.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Joseph, S., & Lewis, C. A. (1995). Factor analysis of the Center for Epidemiological Studies-Depression Scale. *Psychological Reports, 76*, 40–42.

- Kessler, R., Nelson, C. B., McGonagle, K. A., Liu, J., Swartz, M., & Blazer, D. G. (1996). Comorbidity of *DSM-III-R* major depressive disorder in the general population: Results from the US National Comorbidity Survey. *British Journal of Psychiatry*, *168*(Suppl. 30), 17–30.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, *49*, 377–412.
- Radloff, L. S. (1977). The CES-D: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.
- Santor, D. A., & Coyne, J. C. (2001). Evaluating the continuity of symptomatology between depressed and nondepressed individuals. *Journal of Abnormal Psychology*, *110*, 216–225.
- Santor, D. A., Ramsay, J. O., & Zurhoff, D. C. (1994). Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. *Psychological Assessment*, *6*, 255–270.
- Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, *55*, 117–128.
- Watson, D., O'Hara, M. W., Simms, L. J., Kotov, R., Chmielewski, M., & McDade-Montez, E. A. (2006). *Development and validation of the Iowa Depression and Anxiety Scales (IDAS)*. Manuscript submitted for publication.
- Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*, 3–14.

## Prescription Needed for the Ninth Decade of Measuring Depression

Melissa Tanner and David A. F. Haaga  
*Department of Psychology*  
*American University*

In the analysis represented in “Eight Decades in Measurement of Depression,” Santor, Gregus, and Welch (this issue) identify a significant disparity between the development and validation of new measures of the severity of depression and the utilization of such measures, particularly in treatment outcome studies. Although the number of depression measures created has experienced a steady rate of growth since 1960, which has only begun to level off within the past 5 years, the most fre-