Routledge
Taylor & Francis Group

Preliminary Validation of an Online DSM-Based Mental Health Referral Inventory

ROBERT EPSTEIN and LAURA MUZZATTI

University of California San Diego, San Diego, California

A brief online inventory was developed as a much needed corrective for the hundreds of unscientific tests that are used now by millions of people to self-diagnose mental health problems. The primary purpose of the new inventory is to refer people to mental health professionals for further evaluation when they are experiencing problems that might be diagnosable under DSM guidelines; it is not designed to diagnose, however. The inventory was found to be a valid and reliable measuring instrument based on analysis of data obtained from 3,403 subjects. The 54-item checklist looks for 18 common problems identified in the DSM-IV and takes from 5 to 10 min to complete. Test scores proved to be good predictors of a variety of self-reported criterion measures, including happiness, personal and professional success, history of hospitalization, history of therapy, current participation in therapy, employment, and level of education. Females were found to have slightly more mental health problems than males, but no differences in scores were found by race or ethnicity.

KEYWORDS DSM, mental health referrals, mental health test, online testing

The present study consists of a preliminary evaluation of a brief online inventory–the *Epstein Mental Health Inventory* (EMHI)–designed to refer people with possible mental health problems to qualified professionals for further evaluation, available over the Internet free of charge at http://DoYou NeedTherapy.com. Hundreds of psychological tests are available online, but few have been subjected to any validation procedure, and many are contributed by laypeople. A search for the exact phrase "mental disorder test" on Google yields more than 244,000 pages at this writing, suggesting both the proliferation of and demand for such tests. According to a recent report by the Pew Research Center, at least 28% of America's 225 million Internet users—more than 63 million people—have used the Internet to search for information about depression, stress, or other mental health problems, with more people using the Internet for this purpose every year (Fox & Jones, 2009). With so many inadequate tests available and millions of people now relying on the Internet for information about mental health, it is important to make empirically-based tests available and, somehow, to identify them to the public as superior measuring instruments.

Validated comprehensive screening tests for mental health problems exist, but all that we are aware of must be administered by health or mental health professionals, and none are intended for referral purposes. For example, the Holden Psychological Screening Inventory (HPSI; see Aguiar, Reddon, & McNeil, 2003; Holden, Mendonca, Mazmanian, & Reddon, 2006), the Psychological Screening Inventory (PSI; see Lanyon, 2006, 2007), and the Mental Health Inventory 5 (MHI-5; Rumpf, Meyer, Hapke, & Ulrich, 2001) require level-B qualifications for administration or interpretation, which means, at a minimum, completion of graduate level courses in testing or equivalent training. None can be used directly by consumers. Other tests, such as the Structured Clinical Interview for DSM-IV Axis I Disorders (also known as the SCID-1; see Spitzer, Williams, Gibbon, & First, 1992) can be used for diagnosis but must also be administered by clinical professionals. Because it is a comprehensive diagnostic tool, the SCID-1 takes between 45 and 90 min to administer. Another test, called Decisionbase, developed by Canadian psychiatrist Philip W. Long, can also be used for diagnostic purposes and is available online (Kramer & Kennedy, 1998); however, this test appears never to have been scientifically validated.

One widely used test, the 17-item Hamilton Depression Inventory, is in fact validated (e.g., Mottram, Wilson, & Copeland, 2000) and is also currently available online. This inventory screens only for depression, however, rather than for a broad range of mental health problems, and its value even for screening depression has been questioned (Bagby, Ryder, Schuller, & Marshall, 2004). Another validated depression inventory, the Center for Epidemiological Studies Depression Scale (CESD-R), is also available online (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004), as is one version of the

10-item Kessler test (the K10; see Andrew & Slade, 2001; Kessler et al., 2002), which screens only for aspects of depression and anxiety. In addition, the short form of the Drug Abuse Screening Test (DAST; see Carey, Carey, & Chandra, 2003; Skinner, 1982) is available online for screening substance abuse problems. Finally, unofficial, unauthorized, and sometimes inaccurate versions of the Beck Depression Inventory (see Beck, Steer, & Brown, 1996; Cole, Grossman, Prilliman, & Hunsaker, 2003; Osman, Barrios, Gutierrez, Williams, & Bailey, 2008) are also available online; however, the official test is supposed to be obtained only through the authorized publisher (Pearson) and administered by a level C (licensed) mental health professional. The problem, as we see it, is that none of these tests serves as an overall tool for looking at a wide variety of mental health problems that could conceivably be addressed in therapy.

According to the National Institute of Mental Health, more than one in four Americans are suffering from a diagnosable psychiatric disorder at any point in time, but more than two thirds of the population suffering from such problems is never diagnosed or treated, even though effective treatments exist for many disorders (U.S. Department of Health and Human Services, 1999). Because the Internet is now the first place people go for information about their problems (Estabrook, Witt, & Rainie, 2007), validated online tests which screen broadly for mental health problems could become powerful tools for getting people the help they need. We are not suggesting that such tests be used for diagnostic purposes; rather, online screening tests for mental disorders are ideal instruments for prompting people to consult with qualified mental health professionals, who can then determine whether treatment is warranted. The present test was developed with that purpose in mind.

METHODS

Test Design

The present test is distinct from other Internet-based mental health tests in several respects: (a) It includes demographic and criterion questions that can be used for validation purposes; (b) It includes diagnostic criteria from 18 common disorders listed in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR) (American Psychiatric Association, 2000) (Table 1). and, (c) It is fairly comprehensive across mental health problems. The test consists of 54 items in a checklist format.

The test includes three common criteria from each of the 18 major categories of dysfunction, with the 54 items presented in a random order. Technical jargon was avoided (although see Discussion regarding the readability of the items). A typical item was, "Over the past year, my mood has shifted more than once from depressed to highly elevated." Items were

TABLE 1 Prevalence of Disorders by General Population and EMHI Sample

Disorder	General U.S. Population Studies	EMHI (2 Symptoms)	EMHI (3 Symptoms)
Bipolar	0.6–2.6	10.7	13.8
Depression	6.5–10.3	18.3	10.4
Eating	0.1–3.0	3.5	1.1
Generalized Anxiety	2.0–4.0	19.1	19.7
Impulse Control	8.9	14.7	0.8
Mania	1.3	3.8	0.9
OCD	0.5–2.4	14.9	9.8
Other Anxiety	1.6–4.9	9.2	5.6
Other Mood	1.6–6.7	16.4	8.5
Personality	0.5–6.0	7.3	1.9
Phobia	4.0–8.8	10.2	9.2
Psychosis	0.2–0.5	7.0	2.1
PTSD	3.5–3.6	9.3	6.0
Relational*	Unknown	11.8	11.4
Sexual	Unknown**	13.2	1.0
Social Phobia	0.9–6.8	16.6	17.1
Somatoform	0.2–2.0	5.3	0
Substance Abuse	3.8–11.3	6.1	7.2

Sources: Aalto-Setala, Marttunen, Tuulio-Henriksson, Poikolainen, & Lonnqvist, 2001; Ingersoll & Burns, 2001; Kessler, Chiu, Demler, & Walters, 2005; National Institute of Mental Health, 2008; U.S. Department of Health and Human Services, 1999.

*Not currently listed in the DSM-IV, but likely to be included in the DSM-V and to have substantial prevalence (Beach & Kaslow, 2006; Beach, Walboldt et al., 2006; Denton, 2007).

**Although sexual dysfunction is fairly common (Heiman, 2002; Laumann, Paik, & Rosen, 1999), the prevalence of sexual problems that rise to the level of diagnosable disorders is unclear and thought to be quite small (Ingersoll & Burns, 2001).

selected based on the professional judgment of the test developers, with nonoverlapping items that could be easily understood by the average consumer preferred over vague or technical items. Prevalence data are generally unavailable for diagnostic criteria in the DSM and therefore could not be used for guidance in criterion selection (see Discussion). The test taker was directed to check off any of the items which he or she believed to be true.

The 18 categories of dysfunction were: substance abuse, psychosis, depression, mania, bipolar disorder, other mood disorder, phobia, social phobia, obsessive-compulsive disorder, posttraumatic stress disorder, generalized anxiety disorder, other anxiety disorder, relational disorder, sexual disorder, eating disorder, impulse disorder, personality disorder, and somatoform disorder. Although not included in the DSM-IV, relational disorder was included in the present test because of the reasonable possibility that it will be included in the DSM-V and will be found to have high prevalence; it is currently listed in the DSM-IV only as a type of "problem" that is secondary to other disorders (Beach, Wamboldt, et al., 2006; Beach & Kaslow, 2006; Denton, 2007).

Although prevalence data guided the selection of diagnostic categories for this test (Table 1), in order to keep the test short, some DSM disorders with substantial prevalence were not included and may be added to future versions of the test, the main ones being: dysthymia, schizophrenia, panic disorder, agoraphobia, antisocial personality disorder, and unipolar major depression.

In order to assess the possible predictive validity of the test, subjects were asked one question each in seven areas of human functioning that are often affected adversely by mental illness (U.S. Department of Health and Human Services, 1999). Mental illness interferes with both employment and personal relationships, for example, so scores on the EMHI should be able to predict answers to questions on such topics. Typical criterion questions included: "Have you ever been hospitalized for a mental health problem?" (yes or no) and "How much success have you had lately in your personal life?" (10-point scale from low to high). The seven questions covered the following topics: happiness, personal success, professional success, history of hospitalization for mental illness, whether or not the subject was currently employed, whether or not the subject was currently in therapy, and whether or not the subject had ever been in therapy. If the test is a valid instrument, total scores on the test (which correspond to the total number of symptoms one is experiencing) should be negatively correlated with self-reported measures of happiness, employment, personal success, and professional success and should be positively correlated with self-reported measures of hospitalization and therapy. In addition, demographic questions were asked regarding the individual's race/ethnicity, education, and gender.

Upon completion of the test, test takers were given feedback intended to guide those who might be suffering from treatable conditions to consult with a mental health professional; the feedback was based on the number of items that had been checked off in one or more categories—in effect, the number of symptoms one reported. If no items had been checked off, the test taker was complimented on his or her good mental health. If one item had been checked off in any one category, the test taker was advised in mild language that he or she might benefit by consulting with a mental health professional. If two items had been checked off in one or more categories, a stronger recommendation was made. If all three items had been checked off in one or more categories, the test taker was urged to see a mental health professional. If items were checked off in multiple categories, separate recommendations were made. The categories were identified as "areas for possible exploration (expressed in the diagnostic language that will be familiar to your therapist)," and the test taker was also advised that only a qualified professional can diagnose. Links to professional organizations that can put people in touch with therapists were also included.

Sample and Demographics

This study looks at data from 3,403 people who took the test between May 20, 2007, and September 29, 2008. Soon after the test was posted at http://DoYouNeedTherapy.com in May 2007, we also posted links to it at http://DrEpstein.com and http://SelfGrowth.com, and links also began to appear at various independent counseling websites such as http://CounselingPlus.net. It also quickly achieved a relatively high Google search rank using search phrases such as "do you need therapy" and "do I need therapy." No particular steps were taken to advertise the new instrument, so we assume that most people found it when employing search engines to research mental health topics. Although this gave us no control over the makeup of the sample, we were in fact reaching what we viewed as the primary target audience for the test: people using the Internet to evaluate their mental health.

The resulting sample was diverse. Ninety percent of the sample was from the United States and Canada, and the remainder were from 34 other countries. Sixty-one percent were female and 32% male (in addition, eight individuals identified themselves as "other," with 231 unknown). The mean age of test takers was 34.2. Two thousand five hundred and eight identified themselves as White and 660 as non-White (21 American Indian, 277 Asian, 107 Hispanic, 138 Black, and 117 other), with the remainder (235) unknown. Level of education was higher than in the general population (225 none, 1,146 high school, 318 associates degree, 1,152 college, 431 masters, 95 doctorate, 35 unknown).

RESULTS

Validity and Reliability

Test scores proved to be good predictors of the criterion variables (all self-reported), suggesting that the test is a valid measure of mental health problems: happiness (Spearman's $\rho = -0.53^{**}$), personal success ($\rho = -0.42^{**}$), and professional success ($\rho = -0.39^{**}$), as well as employment (Mann-Whitney $U = 872,683^{**}$), history of psychotherapy ($U = 1,153,897^{**}$), hospitalization ($U = 359,884^{**}$), and current participation in therapy ($U = 464,161^{**}$).[1] Mental health problems were also negatively correlated with education level (Kruskal-Wallis $\chi^2 = 127^{**}$). Content validity is assured to some extent because all of the test items were derived from DSM criteria. Symptoms were distributed roughly normally, again suggesting the validity of the measuring instrument (Figure 1); human traits tend to be distributed normally across large populations (Anastasi & Urbina, 2009). Internal-consistency reliability proved to be moderately high using both Cronbach's alpha (0.90) and the Guttman split-half test (0.88).
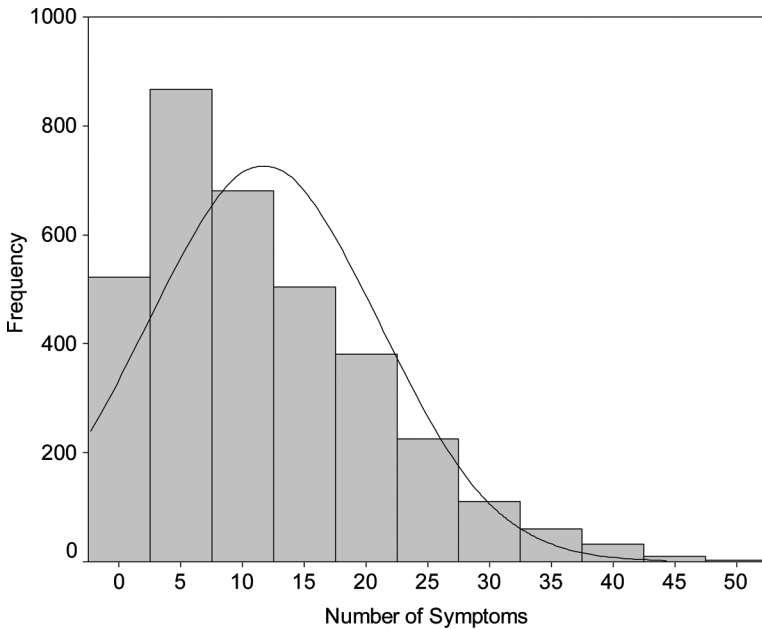
FIGURE 1 Frequency distribution of number of symptoms per subject. A normal curve is superimposed.

Other Results

Gender differences were found, with total scores for females about 17% higher than for males ($\chi^2 = 28.6^{**}$; $\text{mean}_{\text{male}} = 10.6$; $\text{mean}_{\text{female}} = 12.4$). Differences by race and ethnicity on total scores were not found ($\chi^2 = 9.8$, $p = .08\,\text{NS}$). As one might expect, the prevalence of disorders was generally higher than is believed to occur in the general population (Table 1); only 5.8% of test takers reported no symptoms.

DISCUSSION

The use of objective tests for evaluating mental health problems is sometimes called into question; some argue that there is no substitute for the face-to-face clinical interview (e.g., Hunsberger, 2007). Others have argued that an informal interview can be improved through the use of a structured format (Williams et al., 1992) or that a structured interview by a nonclinician can be adequate for assessment (Koenig et al., 1989). We believe that the present test sidesteps the debate about the value of objective testing, at least to some extent. Like it or not, the Internet is being widely used now by millions of people to assess their mental health difficulties (Estabrook et al., 2007; Fox & Jones, 2009). As a validated, DSM-based instrument, the EMHI is almost certainly superior to the assessment tools most Internet users are currently

employing, and its main purpose is in fact to encourage people to visit a therapist, who can then put his or her clinical insights to work. In other words, given that the main function of the EMHI is to prompt Internet users with mental health problems to make contact with qualified clinical professionals, and given that it is not intended for use in assessment or diagnosis, we believe that it fills an important void in the rapidly expanding and chaotic world of Internet testing.

Could a test of this sort harm someone by raising concerns where none are warranted? Although we cannot rule out that possibility, this test has at least undergone some degree of scientific evaluation; an informal test placed online by an untrained individual will almost certainly put people at greater risk. Given the large number of diagnosable people who never receive help for their mental health problems, it can also reasonably be argued that a test that errs on the side of false positives probably does more good than harm. That issue aside, if someone has gone to the trouble of seeking out and taking an online mental health test, he or she probably already has concerns about mental health issues and will likely benefit by consulting with a mental health professional, if only to be reassured.

Subjects in Internet research are necessarily self-selected; in the present study, Internet sampling yielded higher frequencies of symptoms than would be expected in the general population (Table 1). In fact, 94.2% of the test takers were given a referral for at least the lowest level of urgency, and 51.1% of those who were referred were referred at the highest level of urgency. We do not see this as a problem with the test; quite the contrary. The target audience for this type of test is the segment of Internet users who have concerns about their mental health; our data suggest that we are reaching precisely that audience.

Internet sampling also makes it difficult to establish concurrent validity and test–retest reliability. On the positive side, computer screening tools might actually encourage more honest responses than do face-to-face interviews (Barack, 1999; Burke, 1993; Martin & Nagao, 1989). The newest version of the EMHI, not employed in the present study, includes a request for an e-mail address (collected in a way that preserves subject confidentiality), which will allow other kinds of evaluations (e.g., test–retest reliability, concurrent validity, and long-term follow-up) to be conducted in the future. The current version also asks test takers to identify their sexual orientation (in one of five different categories), so future analyses will be able to look at this additional demographic.

As noted earlier, the selection of symptoms specified in the EMHI could not be guided by prevalence data, because prevalence data for DSM criteria generally do not exist; DSM criteria are selected by consensus vote, not by empirical data (Zimmerman, McGlinchey, Young, & Chelminski, 2006). If such prevalence data eventually become available, EMHI criteria can be adjusted accordingly. Of greater concern, because data were collected from anonymous Internet users, concurrent analyses which would have allowed

us to measure the sensitivity and specificity of our subscales could not be conducted. Because this test is meant as a referral tool rather than a diagnostic instrument, it is not clear that the subscales need to be validated in this way; however, because e-mail addresses are now being collected, we anticipate being able to administer multiple tests to test takers in the future and thereby to examine the sensitivity of subscales. Another way to accomplish this would be to have a clinic employ the test during intake for some period of time and then to compare test scores with actual clinical diagnoses.

In addition to determining concurrent validity and test–retest reliability for this instrument, we believe that future versions of the test should probably use simpler language. The Flesch-Kincaid Grade Level score for the EMHI test items is 10.2, which could help explain why the education level of our sample skewed toward higher education. If the language can be made easier to comprehend, the test will likely be helpful for a larger number of people. The challenge is to simplify the language without straying too far from DSM guidelines or undermining the predictiveness of the scores. Because the test is available worldwide, we are also concerned about the way various test items will be understood in cultures outside the United and Canada (10% of the current sample). In the present study, no disclaimers were given regarding the cultural context of the questions and recommendations. Future versions of the test need to include such disclaimers. Ultimately, culturally-sensitive versions of this test, including non-English versions, could be developed that would allow some degree of cross-cultural comparison.

We emphasize that this version of the EMHI is not intended to screen for specific disorders. Rather, its purpose to assess the individual broadly for common mental problems that might benefit from treatment and then to urge the test taker, with various levels of urgency, to seek appropriate help. The fact that total scores on the test are good predictors of several self-reported factors that are related to mental health problems suggests that it has value in this regard. Because the instrument is Internet based, it ultimately could also serve as an efficient gateway to mental health services.

NOTE

1. Nonparametric statistical tests such as Spearman's rho, the Mann-Whitney U, and the Kruskal-Wallis H are used throughout this study because scores on the EMHI lie on an ordinal scale. The double asterisk is used to signify a significance level (p) of less than 0.01. A single asterisk is used to signify a significance level (p) of less than 0.05.

REFERENCES

Aalto-Setala, T., Marttunen, M., Tuulio-Henriksson, A., Poikolainen, K., & Lonnqvist, J. (2001). One month prevalence of depression and other DSM-IV disorders among young adults. *Psychological Medicine*, *31*(5), 791–801.

Aguiar, S., Reddon, J. R., & McNeil, D. C. (2003). Stability and validity of the Holden Psychological Screening Inventory (HPSI) in psychiatric inpatients. *Social Work in Mental Health*, *1*(3), 35–52.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.

Anastasi, A., & Urbina, S. (2009). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Andrew, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale. *Australian and New Zealand Journal of Public Health*, *25*(6), 494–497.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *American Journal of Psychiatry*, *161*, 2163–2177.

Barack, A. (1999). Psychological applications on the Internet: A discipline on the threshold of a new millennium. *Applied and Preventive Psychology*, *8*(4), 231–245.

Beach, S. R. H., & Kaslow, N. J. (2006). Relational disorders and relational processes in diagnostic practice: Introduction to the special section. *Journal of Family Psychology*, *20*(3), 353–355.

Beach, S. R. H., Wamboldt, M. Z., Kaslow, N. J., Heyman, R. E., First, M. B., Underwood, L. G., & Reiss, D. (2006). *Relational processes and DSM-V: Neuroscience assessment, prevention, and treatment*. Arlington, VA: American Psychiatric Publishing.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory–II*. San Antonio, TX: Harcourt Brace.

Burke, M. J. (1993). Computerized psychological testing: Impacts on measuring predictor constructs and future job behavior. In Neal Schmitt, Walter C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 203–239). San Francisco, CA: Jossey-Bass.

Carey, K. B., Carey, M. P., & Chandra, P. S. (2003). Psychometric evaluation of the Alcohol Use Disorders Identification Test and Short Drug Abuse Screening Test with psychiatric patients in India. *Journal of Clinical Psychiatry*, *64*, 767–774.

Cole, J. C., Grossman, I., Prilliman, C., & Hunsaker, E. (2003). Multimethod validation of the Beck Depression Inventory-II and Grossman-Cole Depression Inventory with an inpatient sample. *Psychological Reports*, *93*, 1115–1129.

Denton, W. H. (2007). Issues for DSM-V: Relational diagnosis: An essential component of biopsychosocial assessment. *American Journal of Psychiatry*, *164*, 1146–1147.

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiological Studies Depression Scale: Review and revision (CESD and CESD-R). In Mark Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd ed., pp. 363–377). Mahwah, NJ: Lawrence Erlbaum Associates.

Estabrook, L., Witt, E., & Rainie, L. (2007). *Information searches that solve problems: How people use the Internet, libraries, and government agencies when they need help*. Pew Internet & American Life Project and University of Illinois at

Urbana-Champaign. Retrieved from http://www.pewinternet.org/~/media// Files/Reports/2007/Pew_UI_LibrariesReport.pdf.pdf

Fox, S., & Jones, S. (2009). *The social life of health information*. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf

Heiman, J. R. (2002). Sexual dysfunction: Overview of prevalence, etiological factors, and treatments. *Journal of Sex Research*, *39*(1), 73–78.

Holden, R. R., Mendonca, J. D., Mazmanian, D., & Reddon, J. R. (2006). Clinical construct validity of the Holden Psychological Screening Inventory (HPSI). *Journal of Clinical Psychology*, *48*(5), 627–633.

Hunsberger, P. H. (2007). Reestablishing clinical psychology's subjective core. *American Psychologist*, *62*(6), 614–615.

Ingersoll, R. E., & Burns, L. (2001). *The mental health desk reference: A sourcebook for counselors*. New York, NY: Wiley.

Kessler, R. C., Andrews, G., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*(6), 959–956.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617–627.

Koenig, H. G., Goli, V., Shelp, F., Cohen, H. J., Meador, K. G., & Blazer, D. G. (1989). Major depression and the NIMH Diagnostic Interview Schedule: Validation in medically ill hospitalized patients. *International Journal of Psychiatry in Medicine*, *19*(2), 123–132.

Kramer, T., & Kennedy, R. (1998). Useful websites for psychiatrists. *Academic Psychiatry*, *22*(2), 141–143.

Lanyon, R. I. (2006). Mental health screening: Utility of the Psychological Screening Inventory. *Psychological Services*, *3*(3), 170–180.

Lanyon, R. I. (2007). Utility of the psychological screening inventory: A review. *Journal of Clinical Psychology*, *63*(3), 283–307.

Laumann, E. O., Paik, A., & Rosen, R. C. (1999). Sexual dysfunction in the United States: Prevalence and predictors. *Journal of the American Medical Association*, *281*(6), 537–544.

Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, *74*, 72–80.

Mottram, P., Wilson, K., & Copeland, J. (2000). Validation of the Hamilton Depression Rating Scale and Montgommery and Asberg Rating Scales in terms of AGECAT depression cases. *International Journal of Geriatric Psychiatry*, *15*(12), 1113–1119.

National Institute of Mental Health. (2008). *The numbers count: Mental disorders in America*. Retrieved from http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america.

Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey J. (2008). Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *Journal of Clinical Psychology*, *64*(1), 83–102.

Rumpf, H. J., Meyer, C., Hapke, U., & Ulrich, J. (2001). Screening for mental health: Validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Research*, *105*(3), 243–253.

Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behavior*, 7, 363–367.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III-R (SCID). I: History, rationale, and description. *Archives of General Psychiatry*, *49*(8), 624–629.

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institutes of Mental Health.

Williams, J. B., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., . . . Rounsaville, B. (1992). The structured clinical interview for DSM-III-R (SCID). Multisite test-retest reliability. *Archives of General Psychiatry*, *49*(8), 630–636.

Zimmerman, M., McGlinchey, J. B., Young, D., & Chelminski, I. (2006). Diagnosing major depressive disorder III. Can some symptoms be eliminated from the diagnostic criteria? *Journal of Nervous and Mental Disease*, *194*(5), 313–317.