JOURNAL OF TECHNOLOGY IN HUMAN SERVICES
https://doi.org/10.1080/15228835.2017.1356800

Routledge
Taylor & Francis Group

A DSM-5-Based Online Mental Health Referral Inventory: A Large-Scale Validation Study

Robert Epstein[a], Megan Ho[b], Seojin Hyun[a], Christopher Le[a], Ronald E. Robertson[c], and Dayna Stout[d]

[a]American Institute for Behavioral Research and Technology, Vista, California; [b]Scripps College, Claremont, California; [c]Northeastern University, Boston, Massachusetts; [d]University of California, Berkeley, California

ABSTRACT

In 2007, a comprehensive mental health referral inventory based on the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) was posted online as an alternative to the informal tests the public now uses to self-diagnose mental health problems. A validation study was published in 2011 (Epstein & Muzzatti, 2011). In 2013, a revision of that inventory that was consistent with the DSM-5 was posted. The present study evaluates the new inventory with a diverse sample of 201,625 people from 184 countries (mainly the United States, Canada, and the United Kingdom). The 63-item inventory screens for 21 common problems and typically takes about 5 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, and employment. Females were found to have slightly more mental health problems than males, and effects were also found for education and race/ethnicity. The revised test is more accessible than the original, reducing the reading level from grade 10.2 to 6.6. It is not designed to diagnose; rather, its primary purpose is to refer people to mental health professionals for further evaluation. It can also be used in clinical settings for quick screening purposes.

ARTICLE HISTORY
Received 9 July 2017
Accepted 14 July 2017

KEYWORDS
DSM-5; EMHI; Epstein mental health inventory; mental health screening; mental health tests; online testing

At this writing, a search on the Google search engine for the exact phrase "mental health test" (that phrase with quotation marks around it) yields an impressive 156,000 results. It is easy these days for anyone to cobble together a quiz of some sort and post it online, and the public has no way of distinguishing scientifically validated tests that serve legitimate clinical purposes from informal tests that are mainly for entertainment. First posted online in 2007, the Epstein Mental Health Inventory (EMHI) was closely based on DSM-IV diagnostic criteria and was reviewed by clinical professionals, among them psychiatrist Robert Spitzer, one of the architects of the modern DSM. When people took the test, they were also asked a number of demographic

CONTACT Robert Epstein re@aibrt.org American Institute for Behavioral Research and Technology, 1035 E. Vista Way Ste. 120, Vista, CA 92024.

questions as well as various criterion questions that were later used to evaluate the test. In a study with 3,403 participants in 36 countries (90.0% from the U.S. and Canada), Epstein and Muzzatti (2011) found the test to be both valid (based on the self-reported criterion measures) and to have moderately high internal-consistency reliability.

In May, 2013, the American Psychiatric Association released the DSM-5, its long awaited update to the previous DSM version. Based on advance reports about the DSM-5, efforts had been underway for several months to update the EMHI; the revised inventory (the EMHI-r) was posted online just after the DSM-5 was released. The EMHI had 54 items that screened for 18 disorders. The EMHI-r has 63 items that screen for 21 disorders. One of the main problems with the EMHI was its language; it had a Flesch-Kincaid reading level of 10.2. One of the main purposes of the revision process was to simplify the language while also remaining true to diagnostic criteria.

Epstein and Muzzatti (2011) described a number of existing validated, comprehensive mental health screening tests, such as 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; see Rumpf, Meyer, Hapke, & Ulrich, 2001), noting that "all [the validated tests] that we are aware of must be administered by health or mental health professionals, and none is intended for referral purposes" (Epstein & Muzzatti, 2011, p. 285). They also reported that no validated comprehensive screening tests other than the EMHI appeared to be available online. Other validated tests, such the Hamilton Depression Inventory (Mottram, Wilson, & Copeland, 2000), which was available online in 2011, screened for Depression only. Because some people might not have the presence of mind or the technical knowledge to know which disorder-specific test they should be taking to assess their condition, a comprehensive test is advantageous.

Since 2011, more validated tests have been made available online, but, as far as we are aware, none of them is DSM-5-based, and none offers comprehensive screening. At first glance, a collection of screening tools available at http://www.phqscreeners.com appears to be an exception, but the forms provided there are not self-scorable; they are meant to be completed by patients or clients in a waiting room and then submitted to a clinical professional for analysis. These forms are also based on the DSM-IV. New online, validated tests that can be self-scored include the M3 Screen, a DSM-IV-based test that screens for mood problems (Gaynes et al., 2010) and a number of separate DSM-IV-based tests available at http://MentalHealthAmerica.net, which screen, respectively, for Depression, Anxiety, Bipolar Disorder, PTSD, Psychosis, Youth Problems, and Alcohol and Substance Abuse, each based on validated tests developed by different authors (e.g., Brown & Rounds, 1994; Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, Williams, & Löwe, 2006).

The bigger problem with online tests is that the vast majority of them are non-validated. One website alone—OKCupid.com, which is primarily a dating website—offers more than 45,000 tests, all of which were contributed by users. Even tests that first appear to be reputable can be problematic. At this writing, the first test that turns up when one searches Google for the exact phrase "mental health test" is the Mental Health Assessment offered by *Psychology Today* magazine, accessible at **http://psychologytoday.tests.psychtests.com**. This nonvalidated test screens for eight disorders, and the criteria it uses are DSM-5-based. In some cases, however, it exaggerates the time frames associated with the criteria. For example, for a diagnosis of Generalized Anxiety Disorder, the DSM requires that symptoms have been occurring "more days than not for at least 6 months," but the *Psychology Today* test simply asks, "In the last 6 months, have you experienced any of the following symptoms?" The exaggerated time frames might lead people to seek help unnecessarily. When the user is shown the test results, he or she is also advised, "One useful place to look for licensed therapists is Psychology Today's Therapist Directory," and a link to the directory is shown immediately above. The nature of the test and positioning of the link suggest that the test is serving as a marketing tool for a commercial enterprise.[1]

With Internet penetration now approaching or even exceeding 90% in some countries (Internet Live Stats, 2016), and with more and more people using the Internet to try to diagnose both health and mental health problems using nonvalidated tools that are widely available online (and, increasingly, that are available as mobile apps) (PEW Research Center, 2013), we believe it is important to provide free, scientifically validated online mental health screening tools. The EMHI-r has been developed in that spirit. We also recognize the need to develop tools consistent with the new DSM-5.

Methods

Changes to the test

As noted previously, the EMHI-r screens for 21 disorders, rather than the 18 in the EMHI. Four disorders were removed: Other Anxiety Disorder, Other Mood Disorder, Personality Disorder, and Relational Disorder. And seven disorders were added: Antisocial Personality Disorder, Avoidant Personality Disorder, Borderline Personality Disorder, Dissociative Identity Disorder, Panic Disorder, Persistent Depressive Disorder and Sleep Disorder. As in the previous test, our goal was to include disorders with relatively high prevalence, where prevalence is known (Table 1). Three disorders that might have relatively high prevalence were not included in the new test and might be added to future versions: Attention-Deficit/Hyperactivity Disorder (ADHD), Autism Spectrum Disorder, and Neurocognitive Disorder. Although Relational Disorder had been included in the old test, because of controversies

Table 1. Prevalence of disorders by general population and EMHI-r sample.

Disorder	General U.S. population studies	EMHI-r (2 symptoms)	EMHI-r (3 symptoms)
Antisocial personality disorder	1.0	6.8	3.0
Avoidant personality disorder	5.2	33.4	25.0
Bipolar disorder	2.6	19.7	23.5
Borderline personality disorder	1.6	26.5	9.4
Dissociative identity disorder	1.5	19.1	11.3
Eating disorder	1.2	4.4	1.4
Generalized anxiety disorder	3.1	25.3	33.8
Impulse control disorder	8.9	16.9	11.6
Major depressive disorder	6.7	23.3	40.7
Mania	1.3	7.3	1.3
Obsessive-compulsive disorder	1.0	19.3	15.2
Panic disorder	2.7	17.3	14.2
Persistent depressive disorder	1.5	19.9	21.9
Posttraumatic stress disorder	3.5	17.9	19.5
Psychosis	1.1	9.9	3.8
Sexual disorder	Unknown*	10.8	1.8
Simple phobia	8.7	23.7	20.2
Sleep disorder	28.0	29.8	29.2
Social phobia	6.8	19.2	26.6
Somatic symptom disorder	0.8–5.9	23.0	16.3
Substance abuse disorder	8.2	7.8	5.5

Note. Sources: Centers for Disease Control and Prevention (CDC), 2009; Haller et al., 2015; Kessler et al., 2005; National Institute of Mental Health, 2014; Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, 2014; Johnson, Cohen, Kasen, & Brook, 2006.

*Although sexual dysfunction is not uncommon, sexual disorders are not typically included in large epidemiologic studies (Simons & Carey, 2001).

regarding this diagnostic category (Denton, 2007), it was ultimately included only in an appendix of the DSM-5; hence, we omitted it from the EMHI-r.

Because of the high reading level of the earlier test (Flesch-Kincaid score: 10.2), we took special care to word the items in the new test so that they preserved as much of the meaning of DSM criteria as possible while also being accessible to a wide audience. Thus, one of the items in the Major Depressive Disorder category of the old test read, "For at least the past two weeks, I have found it difficult to get any pleasure from daily activities that I used to enjoy." On the new test, it reads, "For more than two weeks now, it's been hard for me to feel happy doing things I used to enjoy." By rewriting items accordingly, we were able to reduce the Flesch-Kincaid readability score on the new test to 6.6., which suggests that roughly 82% of the U.S. adult population should be able to read and understand it (Murray, Owen, & McGaw, 2005).

As with the previous version of the test, three DSM criteria were used for each of the 21 diagnostic categories included in the EMHI-r. No empirical studies were available to allow to us to determine which of these criteria are the most predictive (Nemeroff et al, 2013; cf. Epstein, 2013). We made our selection based on our best judgment, taking care mainly to include items that overlapped as little as possible and that would be relatively easy for people to understand. In each case, we took pains to eliminate jargon and simplify

language. Thus, one of the criteria used to diagnose Major Depressive Disorder (MDD) in the DSM requires that, "during the same 2-week period," the individual experienced:

> depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful; Note: In children and adolescents, can be irritable mood.) (American Psychiatric Association, 2013, p. 160)

In the EMHI-r, this criterion became, "For more than two weeks now, I've been feeling extremely down and miserable."

For a diagnosis of MDD, the DSM-5 requires a licensed clinical professional to determine that at least five of nine stated criteria are met. Again, for each disorder in the EMHI-r, we listed only three criteria, which means at best that the EMHI-r is a referral tool, not a diagnostic tool, and this limitation is clearly stated both in the introduction to the test and when results are given.

Access and procedure

Like its previous version, the new test was made accessible at http://DoYouNeedTherapy.com. It was also accessible through links at other mental-health-related websites; we had no control over the posting of such links. The introduction to the test stresses that it cannot be used for diagnosis and that "only a qualified, licensed mental health professional" can diagnose. It also indicates that "the test is intended for English speakers only, primarily in the U.S. and Canada." In this regard, it adds that, "If you are not fluent in English, or if you are from some other part of the world, the results might not be valid for you."

Users are then asked a number of demographic and criterion questions and are assured that such information "is being collected for research purposes only and will be kept strictly confidential." Information that could be used to identify individual users, such as e-mail address, full name, or mailing address, was not collected, so anonymity was assured. Demographic questions covered gender, sexual orientation, education, age, and other variables (see as follows), and criterion questions, which should be predictable from test scores, covered such things as history of mental health treatment and hospitalization, diagnostic history, and other variables (see as follows).

Following the demographic and criterion questions, users were instructed as follows: "For each of the following questions, select the answer that best applies to you." Users then clicked on items they felt applied to them. Generally speaking, because people are able to quickly skip over items that don't apply to them, most people take about 5 min. to complete the test.

When finished, people clicked a "submit" button and immediately receive a report, described by Epstein and Muzzatti (2011) as follows:

> 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. (p. 288)

Sample and demographics

The present study examined data obtained from 201,625 people in 184 countries. Before cleaning, 230,790 cases were in the sample; cases were removed when participants answered no questions, when they reported having an English fluency level under 6 (on a 10-point scale), when they appeared to be under age 12 (see as follows), or when they took the test more than once on the same day. When that occurred, we normally kept only the first instance in which they answered both demographic and criterion questions. Because of a technical problem at the company providing our Internet service, we were unable to identify the countries of 66,457 of our participants (33.0% of the 201,625 people in the cleaned sample). Of the remaining 135,168 people, 71,066 were from the United States or Canada (52.6% of the people whose locations were known), 28,975 from the United Kingdom (21.4%), and 35,127 from 181 other countries (26.0%). Data were collected between May 14, 2013 and June 13, 2017.

The mean age of the participants was 22.0 (SD = 9.2, median = 19, mode = 16, range 12 to 94). A minimum-age cutoff of 12, which was approved by our Institutional Review Board, was used given that 12 is the typical age at which students in the United States have completed the 6th grade and given that the reading level of the EMHI-r is 6.6 (although we acknowledge that many people at this age read below their grade level); in any case, only 10,360 (5.3%) of our participants were under age 14.

Fifty two thousand and seventy two (25.8%) of our participants identified themselves as male, 127,877 (63.4%) as female, and 5,906 (2.9%) as other; gender was unknown for 15,770 (7.8%). Forty four thousand, eight hundred and

eighty seven (22.3%) said they lacked a high school degree, 89,492 (44.3%) said they had received a high school degree, 8,960 (4.4%) said they had received an associates degree, 45,273 (22.5%) said they had received a college degree, 9,036 (4.5%) said they had received a masters degree, and 1,733 (0.9%) said they had received a doctorate degree; educational level was unknown for 2,334 (1.2%).

One hundred thirty five thousand and one hundred and two (67.0%) identified themselves as White, 19,339 (9.6%) as Asian, 12,414 (6.2%) as Hispanic, 9,095 (4.5%) as Black, 1,228 (0.6%) as American Indian, and 13,810 (6.8%) as other; race or ethnicity was unknown for 10,637 (5.3%). One hundred eight thousand and three hundred and thirty (53.7%) identified themselves as straight, 21,557 (10.7%) as bisexual, 19,271 (9.6%) as unsure, 8,716 (4.3%) as other, and 8,377 (4.2%) as gay; sexual orientation was unknown for 35,374 (17.5%). Sixty three thousand and five hundred and ninety three (31.5%) identified themselves as employed and 97,840 (48.5%) as unemployed; employment status was unknown for 40,192 (19.9%). Finally, the average English fluency reported by the people included in our study was 9.4 (SD = 1.0).

Criterion questions

Regarding questions we asked to assess the predictive validity of the test: 64,906 (32.2%) of our participants reported they had been in therapy, and 135,852 (67.4%) said they had not; history of therapy ever was unknown for 867 (0.4%). Fourteen thousand and nine hundred and eighty nine (7.4%) reported they had been hospitalized for a mental problem, and 185,579 (92.0%) said they had not; hospitalization was unknown for 1,057 (0.5%). Fourteen thousand and six hundred and sixty seven (7.3%) reported they were currently in therapy, and 185,938 (92.2%) said they were not; therapy status was unknown for 1,020 (0.5%). In addition, on a scale of 1 to 10 where 1 was low and 10 was high, when our participants were asked how "happy and fulfilled" they were, they reported an average score of 4.1 (SD = 2.0). When asked how much success they had had lately in their personal life, they reported an average score of 3.9 (SD = 2.3). When asked how much success they had had lately in their professional life, they reported an average score of 4.3 (SD = 2.6).

Results

Validity and reliability

As with the previous version of the test, overall test scores (a sum of the number of symptoms that were checked off) proved to be fairly good predictors of self-reported criterion measures, which suggests that the EMHI-r is a valid

instrument for measuring mental health problems. Mean total scores were negatively correlated with self-reported happiness (Spearman's $\rho = -0.42^{***}$), personal success ($\rho = -0.31^{***}$), and professional success ($\rho = -0.29^{***}$), and they were also good predictors of unemployment (Mann-Whitney $U = 2{,}461{,}456{,}559^{***}$, $M_{\text{unemployed}} = 38.5$, $M_{\text{employed}} = 31.6$), history of therapy ($U = 3{,}760{,}019{,}768^{***}$, $M_{\text{yes}} = 39.93$, $M_{\text{no}} = 34.3$), history of hospitalization ($U = 959{,}012{,}429.5^{***}$, $M_{\text{yes}} = 46.2$, $M_{\text{no}} = 35.1$), and whether participants were currently in therapy ($U = 1{,}069{,}014{,}574^{***}$, $M_{\text{yes}} = 42.9$, $M_{\text{no}} = 35.4$).[2] Although content validity was not assessed directly, it is assured to some extent give that all of the items on the EMHI-r are based on DSM-5 diagnostic criteria. Test–retest reliability could also not be assessed, but internal-consistency reliability was fairly high (Cronbach's alpha = 0.93; Guttman split-half = 0.93).

Other results

Mean total scores differed significantly by educational level (Kruskal-Wallis $H = 9908.37^{***}$, $M_{\text{none}} = 40.0$, $M_{\text{high school}} = 37.8$, $M_{\text{associates}} = 32.3$, $M_{\text{college}} = 31.5$, $M_{\text{masters}} = 24.7$, $M_{\text{doctorate}} = 22.2$), and total scores were negatively correlated with age ($\rho = -0.22^{***}$). An effect was also found for gender ($H = 3320.53^{***}$, $M_{\text{Male}} = 32.6$, $M_{\text{Female}} = 36.9$, $M_{\text{Other}} = 45.0$), with those identifying as "other" scoring higher than those identifying as male or female (cf. Warren, Smalley, & Barefoot, 2016). An effect was also found for race and ethnicity ($H = 195.93^{***}$, $M_{\text{American Indian}} = 39.1$, $M_{\text{Asian}} = 34.9$, $M_{\text{Black}} = 34.3$, $M_{\text{Hispanic}} = 36.7$, $M_{\text{Other}} = 36.7$, $M_{\text{White}} = 36.1$), with American Indians scoring higher than other groups (cf. Beals et al., 2005; Walls, Johnson, Whitbeck, & Hoyt, 2006).

An effect was also found for sexual orientation ($H = 5848.26^{***}$, $M_{\text{bisexual}} = 41.6$, $M_{\text{gay/lesbian}} = 38.5$, $M_{\text{other}} = 42.9$, $M_{\text{straight}} = 33.2$, $M_{\text{unsure}} = 39.3$), with self-labeled straights scoring lower than other groups, consistent with the findings of other studies (e.g., Jorm, Korten, Rodgers, Jacomb, & Christensen, 2002). An effect was also found for country ($H = 751.18^{***}$, $M_{\text{UnitedStates/Canada}} = 34.6$, $M_{\text{UnitedKingdom}} = 38.1$, $M_{\text{Other}} = 36.1$).

As with the EMHI, total scores were roughly normally distributed (Figure 1), which suggests that the EMHI-r is a valid measuring instrument; across large populations, human traits tend to be normally distributed (Anastasi & Urbina, 2009).

Finally, and consistent with the findings of the 2010 National Comorbidity Study (Merikangas et al., 2010), we found mental health problems to be more common among minors than among adults. The difference between mean total scores was substantial, and so were the differences between mean scores in all 21 of the subscales (Table 2); minors outscored adults in 19 of the disorders, and adults outscored minors in two categories (Substance Abuse and Sexual Disorders).

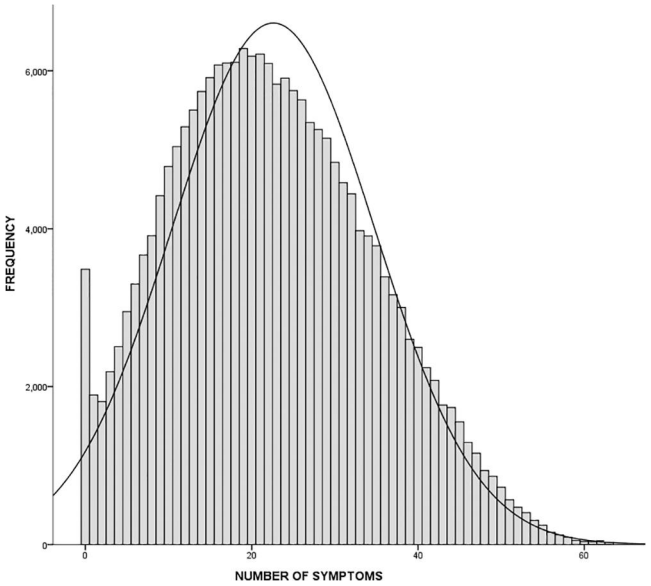

Figure 1. Frequency distribution of number of symptoms per participant. A normal curve is superimposed. The spike at 0 is caused by the large number of people who did not check off any symptoms. As long as they provided demographic information and answered criterion questions, we viewed them as legitimate test takers and kept them in the study.

Table 2. Mental disorder differences between adults and minors.

	Mean Symptoms		
Mental disorder	Under 18 ($N = 75{,}955$)	Over 18 ($N = 120{,}231$)	Difference
Total	40.11	33.21	−6.90***
Antisocial personality disorder	0.48	0.34	−0.14***
Avoidant personality disorder	1.80	1.59	−0.21***
Bipolar disorder	1.52	1.25	−0.27***
Borderline personality disorder	1.44	1.15	−0.29***
Dissociative identity disorder	1.23	0.87	−0.36***
Eating disorder	0.36	0.18	−0.18***
Generalized anxiety disorder	1.80	1.68	−0.12***
Impulse control disorder	1.06	0.87	−0.19***
Major depressive disorder	2.01	1.75	−0.26***
Mania	0.57	0.34	−0.23***
Obsessive–compulsive disorder	1.38	0.99	−0.39***
Panic disorder	1.17	0.83	−0.34***
Persistent depressive disorder	1.33	1.23	−0.10***
Posttraumatic stress disorder	1.21	1.07	−0.14***
Psychosis	0.81	0.46	−0.35***
Sexual disorder	0.31	0.57	0.26***
Simple phobia	1.49	1.25	−0.24***
Sleep disorder	1.86	1.58	−0.28***
Social phobia	1.70	1.21	−0.49***
Somatic symptom disorder	1.51	1.14	−0.37***
Substance abuse disorder	0.26	0.58	0.32***

***$p < 0.001$. Significance was assessed using the Mann-Whitney U test.

Regressions and factor analysis

Linear regression was used to determine which of the 21 disorders included in the EMHI-r best predicted various self-reported outcome measures. For the three criterion variables that were on 10-point scales—happiness, personal success, and professional success—the best predictors were Major Depressive Disorder (for happiness: $\beta = -0.42^{***}$; for personal success: $\beta = -0.31^{***}$; for professional success: $\beta = -0.19^{***}$) and Persistent Depressive Disorder (for happiness: $\beta = -0.22^{***}$; for personal success: $\beta = -0.15^{***}$; for professional success: $\beta = -0.19^{***}$).

An exploratory principal components factor analysis was performed that included all 63 items. The appropriateness of our data for factor analysis was confirmed by a high Kaiser-Meyer-Olkin measure of sampling adequacy (0.95) and a significant Bartlett's test of sphericity ($P < 0.001$). Overall, the analysis yielded six distinct, interpretable, and statistically sound components (Table 3): (a) persistent fear, (b) depression and mood problems, (c) impulsivity and instability, (d) feelings of inadequacy, (e) sex and drug problems, and (f) poor self-care.

Discussion

Epstein and Muzzatti (2011) argued that no matter what the validity of objective tests such as the EMHI, we now live in a world in which many if not most people turn first to the Internet to get information about health and mental health, often by using nonvalidated tests and quizzes (to explore these issues in detail, see Burns, Davenport, Durkin, Luscombe, & Hickie, 2010; Horgan & Sweeney, 2010; Lam-Po-Tang & McKay, 2010; PEW Research Center, 2013). With Internet penetration increasing substantially every year, that is even truer today than it was in 2011, and this trend will likely continue for the foreseeable future. With only a handful of validated measuring instruments available online, we believe the EMHI-r has special value, at least in part because it is, to our knowledge, the only validated, comprehensive, DSM-5-based mental health test currently available online. The main flaw in the previous version—the high Flesch-Kincaid score (10.2)—has been remedied in the new version (reading level 6.6), and, given the relatively high validity measures we found in the present study (comparable in all respects to those reported by Epstein and Muzzatti (2011)), we believe the readability problem has been solved with little or no loss of substance.

Because the present study was Internet-based, it suffers from the usual disadvantages of such studies, the main one being a lack of control over the sample (cf. Birnbaum, 2004; Hewson, 2003). Of particular concern, the mean age of the participants in the present study was only 22.0 (median = 19), whereas the mean age in the Epstein and Muzzatti (2011) study was 34.2

Table 3. Factor loadings for the test items.

Item	Persistent fear	Depression and mood problems	Impulsivity and instability	Feelings of inadequacy	Sex and drug problems	Poor self-care
43	.620					
46	.616					
63	.581					
7	.573					
51	.570					
2	.564					
31	.556					
45	.541					
18	.522					
14	.475					
48	.439					
5	.427					
61	.409					
15		.623				
22		.595				
10		.552				
54		.528				
62		.517				
38		.495				
40		.493				
30		.492				
58		.491				
6		.472				
1		.468				
41		.466	.452			
24		.438				
60			.587			
37			.548			
9			.542			
42			.531			
32			.520			
12			.514			
55			.493			
11			.484			
13			.481			
34		.407	.460			
8			.433			
16				.676		
4				.610		
36				.597		
56				.562		
23				.496		
53				.487		
57					.680	
20					.676	
29					.661	
21					.491	
44					.458	
35						.681
49						.649
27						.502
28						.429

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Factor loadings under 0.40 are not shown.

(the median was not reported). The median age in the United States was recently estimated to be 37.9 (Central Intelligence Agency [CIA], 2013); worldwide, as of 2010, it was estimated to 29 (United Nations Department of Economic and Social Affairs, 2014).

Also of concern, scores on the EMHI-r are likely inflated because people who are drawn to take an online mental health test are presumably concerned about their mental health. In fact, 99.2% of our test takers were referred to a mental health professional at the lowest level of urgency, and 78.3% were referred at the highest level of urgency—far more than one would expect in the general population (Haller, Cramer, Lauche, & Dobos, 2015; Kessler, Chiu, Demler, & Walters, 2005). Because people took the test anonymously online, we were also unable to assess two important measures of test value: test–retest reliability and concurrent validity.

That said, the present study also benefitted from two of the remarkable advantages of conducting research over the Internet: Our sample was extremely large and diverse (201,625 participants from 184 countries), for one thing, and a growing body of research suggests that anonymous Internet-based surveys tend to elicit more honest answers from people than do traditional survey methodologies, especially when people are being questioned about socially sensitive topics such as mental illness (Bharadwaj, Pai, & Suziedelyte, 2015; Conner et al., 2010; Corrigan, 2004; Manderscheid et al., 2010; Robertson, Tran, Lewark, & Epstein, 2017).

Regarding the potential harm that the EMHI-r might do, we quote Epstein and Muzzatti (2011):

> 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. (p. 291)

Finally, we emphasize that the EMHI-r should not be used for diagnostic purposes. We believe, however, that it serves an important purpose on the Internet and that it might also be useful as a waiting room survey in clinical settings.

Recommendations for future research

As noted previously, empirical studies that rate the validity of specific DSM diagnostic criteria generally do not exist, which forced us to use our best

judgment in selecting criteria for the survey instrument. As relevant studies are published, we intend to improve the instrument accordingly. Adding other criteria from the DSM and conducting item analyses to determine which criteria are the best predictors of various outcome measures might also be helpful.

The diagnostic categories included in the inventory also need to be reviewed. As noted earlier, three fairly common disorders—Attention-Deficit/Hyperactivity Disorder (ADHD), Autism Spectrum Disorder, and Neurocognitive Disorder—were not included in the EMHI-r but might be included in a revision. ADHD was excluded in part because of a long-running dispute about the validity of the diagnostic category (e.g., see Saul, 2014) but also because it is diagnosed more frequently in children than adults (Kessler et al., 2005; Rowland et al., 2015). Autism Spectrum Disorder and Neurocognitive Disorder were excluded out of concerns that the test taker might have difficulty understanding some of the content; the same concerns could be raised, however, about Psychosis, Substance Abuse, and other disorders that were included. The solution to this problem, we believe, is to create another version of the inventory that parents, teachers, or even mental health professionals could use to evaluate the symptomology of people in their care. The most important follow-up research that needs to be conducted, however, is not about the EMHI-r itself. The Internet is not only making good information available to people; it is also making vast volumes of bad information available, and people have no easy way of distinguishing the good from the bad. A critically important area of investigation, we believe, is to determine how to make sure that people who are worried about health or mental health issues are quickly referred to accurate information and valid assessment tools. No matter how strong the numbers, the EMHI-r is useless unless (a) people can find it and (b) they trust its results more than the results they might obtain from any of the thousands of nonvalidated tests that now flood cyberspace.

Acknowledgments

Based on a paper presented at the 95th annual meeting of the Western Psychological Association, Las Vegas, April 2015. We thank Jessica Aceret, Nicholas Albrecht, Jessica Antt, Carlos Garcia, Katrina Kaminaka, Laura Muzzatti, and Jennifer Ong for help in developing and updating the EMHI.

Notes

1. Full disclosure: The first author held various editorial positions at *Psychology Today* between 1996 and 2006. He currently has no affiliation with the magazine.
2. 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-r lie on

an ordinal scale. A single asterisk is used to signify a significance level (p) of less than 0.05. A double asterisk is used to signify a significance level (p) of less than 0.01. A triple asterisk is used to signify a significance level (p) of less than 0.001.

References

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. doi:10.1300/j200v01n03_03

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

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

Beals, J., Novins, D. K., Whitesell, N. R., Spicer, P., Mitchell, C. M., & Manson, S. M. (2005). Prevalence of mental disorders and utilization of mental health services in two American Indian reservation populations: Mental health disparities in a national context. *American Journal of Psychiatry*, *162*(9), 1723–1732. doi:10.1176/appi.ajp.162.9.1723

Bharadwaj, P., Pai, M. M., & Suziedelyte, A. (2015). *Mental health stigma*. Retrieved from http://voxeu.org/article/mental-health-stigma

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Psychology*, *55*(1), 803.

Brown, R. L., & Rounds, L. A. (1994). Conjoint screening questionnaires for alcohol and other drug abuse: Criterion validity in a primary care practice. *Wisconsin Medical Journal*, *94*(3), 135–140.

Burns, J. M., Davenport, T. A., Durkin, L. A., Luscombe, G. M., & Hickie, I. B. (2010). The Internet as a setting for mental health service utilization by young people. *Medical Journal of Australia*, *192*(11), S22.

Centers for Disease Control and Prevention (CDC). (2009). Perceived insufficient rest or sleep among adults-United States, 2008. *Morbidity and mortality weekly report*, *58*(42), 1175.

Central Intelligence Agency (CIA). (2013). *The world factbook 2013-14*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/fields/2177.html

Conner, K. O., Copeland, V. C., Grote, N. K., Koeske, G., Rosen, D., Reynolds, C. F., & Brown, C. (2010). Mental health treatment seeking among older adults with depression: The impact of stigma and race. *The American Journal of Geriatric Psychiatry*, *18*(6), 531–543. doi:10.1097/jgp.0b013e3181cc0366

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7), 614.

Denton, W. H. (2007). Issues for DSM-V: Relational diagnosis: An essential component of biopsychosocial assessment. *American Journal of Psychiatry*, *164*(8), 1146–1147. doi:10.1176/appi.ajp.2007.07010181

Epstein, R. (2013, September/October). Woe is us: A review of *The Book of Woe* by Gary Greenberg. *Scientific American Mind*, 70.

Epstein, R., & Muzzatti, L. (2011). Preliminary validation of an online DSM-based mental health referral inventory. *Journal of Technology in Human Services*, *29*, 284–295. doi:10.1080/15228835.2011.638421

Gaynes, B. N., DeVeaugh-Geiss, J., Weir, S., Gu, H., MacPherson, C., Schulberg, H. C., Rubinow, D. R. (2010). Feasibility and diagnostic validity of the M-3 checklist: A brief, self-rated screen for depressive, bipolar, anxiety, and post-traumatic stress disorders in primary care. *Annals of Family Medicine*, *8*(2), 160–169. doi:10.1370/afm.1092

Haller, H., Cramer, H., Lauche, R., & Dobos, G. (2015). Somatoform disorders and medically unexplained symptoms in primary care. *Dtsch Arztebl Int*, *112*, 279–287.

Hewson, C. (2003). *Internet research methods: A practical guide for the social and behavioural sciences*. London, United Kingdom: Sage Publications.

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. doi:10.1002/1097-4679(199209)48:5<627::aid-jclp2270480508>3.0.co;2-9

Horgan, A., & Sweeney, J. (2010). Young students' use of the Internet for mental health information and support. *Journal of Psychiatric and Mental Health Nursing*, *17*(2), 117–123. doi:10.1111/j.1365-2850.2009.01497.x

Internet Live Stats. (2016). *Internet users by country (2016)*. Retrieved from http://www.internetlivestats.com/internet-users-by-country/

Johnson, J. G., Cohen, P., Kasen, S., & Brook, J. S. (2006). Dissociative disorders among adults in the community, impaired functioning, and axis I and II comorbidity. *Journal of Psychiatric Research*, *40*(2), 131–140.

Jorm, A. F., Korten, A. E., Rodgers, B., Jacomb, P. A., & Christensen, H. (2002). Sexual orientation and mental health: Results from a community survey of young and middle-aged adults. *The British Journal of Psychiatry*, *180*(5), 423–427. doi:10.1192/bjp.180.5.423

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 593–602. doi:10.1001/archpsyc.62.6.593

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.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of General Internal Medicine*, *16*(9), 606–613. doi:10.1046/j.1525-1497.2001.016009606.x

Lam-Po-Tang, J., & McKay, D. (2010). Dr Google, MD: A survey of mental health-related Internet use in a private practice sample. *Australasian Psychiatry*, *18*(2), 130–133. doi:10.3109/10398560903473645

Lanyon, R. I. (2006). Mental health screening: Utility of the psychological screening inventory. *Psychological Services*, *3*(3), 170–180. doi:10.1037/1541-1559.3.3.170

Lanyon, R. I. (2007). Utility of the psychological screening inventory: A review. *Journal of Clinical Psychology*, *63*(3), 283–307. doi:10.1002/jclp.20344

Manderscheid, R., Delvecchio, P., Marshall, C., Palpant, R. G., Bigham, J., Bornemann, T. H., … Lubar, D. (2010). Attitudes toward mental illness-35 states, District of Columbia, and Puerto Rico, 2007. *Morbidity and Mortality Weekly Report*, *59*(20), 619–625.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989. doi:10.1016/j.jaac.2010.05.017

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. doi:10.1002/1099-1166(200012)15:12<1113::aid-gps253>3.0.co;2-y

Murray, T. S., Owen, E., & McGaw, B. (2005). *Learning a living: First results from adult literacy and lifeskills survey*. Paris, France: Statistics Canada and Organization for Economic Co-operation and Development.

National Institute of Mental Health. (2014). *What is prevalence?* Retrieved from https://www.nimh.nih.gov/health/statistics/prevalence/index.shtml

Nemeroff, C. B., Weinberger, D., Rutter, M., MacMillan, H. L., Bryant, R. A., Wessely, S., … Malhi, G. S. (2013). DSM-5: A collection of psychiatrist views on the changes, controversies, and future directions. *BMC medicine*, *11*(1), 1.

PEW Research Center. (2013). *The Internet and health*. Retrieved from http://www.pewinternet.org/2013/02/12/the-internet-and-health/

Robertson, R. E., Tran, F. W., Lewark, L. N., & Epstein, R. (2017). Estimates of non-heterosexual prevalence: The roles of anonymity and privacy in survey methodology. *Archives of Sexual Behavior*. doi:10.1007/s10508-017-1044-z

Rowland, A. S., Skipper, B. J., Umbach, D. M., Rabiner, D. L., Campbell, R. A., Naftel, A. J., & Sandler, D. P. (2015). The prevalence of ADHD in a population-based sample. *Journal of Attention Disorders*, *19*(9), 741–754. doi:10.1177/1087054713513799

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. doi:10.1016/s0165-1781(01)00329-8

Saul, R. (2014). *ADHD does not exist: The truth about attention deficit and hyperactivity disorder*. New York, NY: HarperCollins.

Simons, J., & Carey, M. P. (2001). Prevalence of sexual dysfunctions: Results from a decade of research. *Archives of Sexual Behavior*, *30*(2), 177–219.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. (2014) *The NSDUH Report: Substance Use and Mental Health Estimates from the 2013 National Survey on Drug Use and Health: Overview of Findings*. Rockville, Maryland: Retrieved from http://www.samhsa.gov/data/sites/default/files/NSDUH-SR200-RecoveryMonth-2014/NSDUH-SR200-RecoveryMonth-2014.pdf

United Nations Department of Economic and Social Affairs. (2014). *World population ageing 2013*. Statistical Papers–United Nations (Ser. A), Population and Vital Statistics Report, pp. 1–114. doi:10.18356/30d0966c-en

Walls, M. L., Johnson, K. D., Whitbeck, L. B., & Hoyt, D. R. (2006). Mental health and substance abuse services preferences among American Indian people of the northern Midwest. *Community Mental Health Journal*, *42*(6), 521–535. doi:10.1007/s10597-006-9054-7

Warren, J. C., Smalley, K. B., & Barefoot, K. N. (2016). Psychological well-being among transgender and genderqueer individuals. *International Journal of Transgenderism*, *17*(3–4), 114–123. doi:10.1080/15532739.2016.1216344