

A Comparison of Three Brief Depression Measures in an Acute Psychiatric Population: CES-D-10, QIDS-SR, and DASS-21-DEP

Rachel B. Weiss · Idan M. Aderka · Josephine Lee · Courtney Beard · Thröstur Björgvinsson

Published online: 30 September 2014
© Springer Science+Business Media New York 2014

Abstract The CES-D-10, QIDS-SR, and DASS-21-DEP are brief self-report instruments for depression that have demonstrated strong psychometric properties in clinical and community samples. However, it is unclear whether any of the three instruments is superior for assessing depression and treatment response in an acute, diagnostically heterogeneous, treatment-seeking psychiatric population. The present study examined the relative psychometric properties of these instruments in order to inform selection of an optimal depression measure in 377 patients enrolled in a psychiatric partial hospital program. Results indicated that the three measures demonstrated good to excellent internal consistency and strong convergent validity. They also demonstrated fair to good diagnostic utility, although diagnostic cut-off scores were generally higher than in previous samples. The three measures also evidenced high sensitivity to change in depressive symptoms over treatment, with the QIDS-SR showing the strongest effect. The results of this study indicate that any of the three depression measures may satisfactorily assess depressive symptoms in an acute psychiatric population. Thus, selection of a specific assessment tool should be guided by the identified purpose of the assessment. In a partial hospital setting, the QIDS-SR may confer some advantages, such as correspondence with DSM criteria, greater sensitivity to change, and assessment of suicidality.

Keywords Depression · Assessment · Psychometric · Center for the epidemiological studies of depression-short form · Quick inventory of depressive symptomatology-self-report version · Depression anxiety and stress scales-21-item version

The identification of valid, clinically useful, and efficient assessment tools is particularly important in the context of increasing emphasis on evidence-based mental healthcare and outcomes evaluation (Hunsley and Mash 2005). Yet, far less attention has been paid to issues related to the evidence-based assessment of depression in real-world clinical settings, relative to the emphasis placed on evidence-based treatment approaches in these settings (Barlow 2005; Hunsley and Mash 2005). Although there are a number of instruments available (see Joiner et al. 2005), there is a dearth of empirically-based information to guide the selection of depression measures in acute, heterogeneous treatment populations. Assessing treatment progress and outcome in acute treatment settings (e.g., inpatient, residential, or partial hospitals) is just as critical as other settings and warrants its own empirically-based evidence given the unique characteristics of these settings (e.g., very limited time to provide treatment in a population with high levels of symptom severity, comorbidity, functional impairment, and suicide risk).

Although the clinical benefits of using evidence based assessments to monitor treatment outcome are clear (e.g., Duffy et al. 2008; Slade et al. 2006), they remain underutilized in psychiatric settings (Weiss et al. 2009; Zimmerman and McGlinchey 2008; Gilbody et al. 2002). This is in part due to a lack of recognition of the clinical benefits, as well as practical barriers such as time (Zimmerman and McGlinchey 2008). There are unique challenges faced by acute psychiatric settings, such as multiple presenting problems in individual patients and across patients. Moreover, these settings are often

R. B. Weiss (✉) · I. M. Aderka · J. Lee · C. Beard · T. Björgvinsson
McLean Hospital/Harvard Medical School, 115 Mill St., Belmont,
MA 02478, USA
e-mail: rbweiss@mclean.harvard.edu

I. M. Aderka
Department of Psychology, University of Haifa, Mt. Carmel 31905,
Israel

characterized by a fast pace, high census, and brief treatment durations. For example, the partial hospital in which the current study took place admits approximately 18 patients per week with lengths of stay typically between 3 and 10 days. Thus, symptom-specific scales need to be especially short, user-friendly, easy to score and interpret, and provide meaningful data that the clinicians would not otherwise have obtained. Although many instruments may fit these requirements, there is currently little information to guide the selection of a depression measure in acute settings. The psychiatric hospital at which the current study was conducted requires that all units conduct progress monitoring, and several measures are offered as potential assessments. However, we were unable to find any studies comparing these measures to offer guidance on which measure would best suit our clinical needs. Thus, the aim of the present study was to compare the psychometric properties of depression measures in an acute, diagnostically complex psychiatric population.

We selected three measures, all of which addressed barriers to assessment in acute, psychiatric settings. Specifically, we selected measures that have a self-report format, are brief, and available in the public domain. Additionally, we selected two of the measures (i.e., Center for Epidemiologic Studies Depression Scale and Quick Inventory of Depressive Symptomatology - Self-report version) because they are offered as potential outcome measures in our hospital's ongoing clinical measurement initiative, perhaps making these measures more likely to be used by other psychiatric settings. We selected short versions of these measures when available, as brief measures are more feasible to implement in real world settings with multiple competing demands on clinician time and patients who are more easily fatigued and distressed.

The first instrument evaluated in the present study was a ten-item version of the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff 1977). The CES-D was primarily designed for use in studies of the epidemiology of depressive symptoms in the general population. The original version of the CES-D consisted of 20 items and evidenced strong psychometric properties in assessing depressive symptoms and detecting a depression diagnosis. Based on item-total correlations, a shortened ten-item version (CES-D-10) was developed that offers improved efficiency and ease of scoring (Andresen et al. 1994). In that study of healthy older adults, the CES-D-10 demonstrated good test-retest reliability ($r=.71$) and adequate item-total correlations ($r_s=.45-.71$). The short version also showed positive correlations with related constructs such as stress ($r=.43$) and expected negative correlations with positive affect ($r=-.63$; Andresen et al. 1994). As that study examined a healthy population, no information was provided about diagnostic or treatment sensitivity.

The CES-D-10 has been validated for use in multi-cultural elderly populations (Boey 1999; Cheng and Chan 2005; Cheng et al. 2006; Lee and Chokkanathan 2008) and in adolescents (Bradley et al. 2010). It has also been shown to adequately screen for suicidal thoughts and behaviors in a Chinese community sample (Cheung et al. 2007). The CES-D-10 has demonstrated promising psychometric properties as a measure of overall depressive symptomatology and has been found to have a single underlying factor in a sample from the partial hospital in the present study (Björgevinnsson et al. 2013). Although the 20-item CES-D has been noted to be one of the best available screening tools for depression (Joiner et al. 2005), research has yet to examine the CES-D-10 as an outcome measure of depressive symptoms in an acute psychiatric population or how it compares to other available measures.

The Quick Inventory of Depressive Symptomatology - Self-report version (QIDS-SR; Rush et al. 2003; Trivedi et al. 2004) is a 16-item questionnaire that was designed to efficiently measure depressive symptom severity and assess treatment outcomes. The instrument was developed among clinical samples, and assesses the nine criterion symptom domains that define a major depressive episode according to the Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV-TR; American Psychiatric Association 2000). The QIDSSR is a shortened form of the 30-item Inventory of Depressive Symptomatology-Self report version (Rush et al. 1996, 2000), and it has demonstrated equal or better sensitivity to symptom change compared to its longer counterpart (Rush et al. 2003; 2005, 2006b; Trivedi et al. 2004). The QIDS-SR has also performed well against a matched clinician-rated version (Rush et al. 2006a, b; Trivedi et al. 2004), and against 17- and 24-item versions of the clinician-rated Hamilton Rating Scale for Depression (Rush et al. 2003; 2005, a). Psychometric evaluations in depressed outpatients have demonstrated that the QIDS-SR is a unidimensional questionnaire with high internal consistency ($\alpha=.87$), good construct validity (correlation with Hamilton Rating Scale for Depression=.86), and sensitivity to treatment effects in outpatients with major depressive disorder (Rush et al. 2006a; for similar psychometric properties, see also Rush et al. 2003; 2005; Trivedi et al. 2004). However, to our knowledge the QIDS-SR has not been evaluated in acute patients in non-outpatient settings.

The Depression, Anxiety, and Stress Scales are available in both 42- and 21-item versions (DASS and DASS-21, respectively; Lovibond and Lovibond 1995a, b). The DASS-21 is designed to assess and provide maximum discrimination between the core symptoms of depression, anxiety, and overall stress. Thus, the DASS-21 is appealing for assessment in heterogeneous populations, as it can provide information about three related, but distinct symptom domains. The instrument was developed on an empirical basis from non-clinical

populations (Lovibond and Lovibond 1995a, b). The depression subscale of the DASS-21 (DASS-21-DEP) is comprised of seven items, inquiring about symptoms typically associated with dysphoric mood. In a sample of medical patients seeking treatment for worry, the DASS-21-DEP demonstrated strong psychometric properties, including internal consistency ($\alpha=.87$), convergent validity (e.g., correlation with BDI-II, $r=.76$), discriminative validity (e.g., correlation with quality of life $r=-.58$), divergent validity (participants with a mood disorder scored higher than those without), and diagnostic utility (equivalent to BDI-II in ability to detect a diagnosis of MDD; Gloster et al. 2008). Similarly strong psychometric properties have been found in several other samples such as mood and anxiety outpatients (Antony et al. 1998), undergraduates (Norton 2007), general adult (Henry and Crawford 2005), and pain patients (Wood et al. 2010).

Some studies have supported the hypothesized three-factor structure of the overall DASS-21 (Antony et al. 1998; Gloster et al. 2008; Norton 2007; Wood et al. 2010), whereas others have not (Henry and Crawford 2005; Patrick et al. 2010). However, Ronk and colleagues (2014) recently found that the hypothesized three-factor structure of the DASS-21 had excellent clinical utility and sensitivity to treatment in a large sample ($n=3964$) of inpatients. Similarly, Ng et al. (2007) evaluated the validity of the DASS-21 as a routine clinical outcome measure in an inpatient setting and found that the instrument was reliable, valid, and sensitive to change in treatment. Thus, the DASS-21 may perform well in other acute settings, such as partial hospital.

In sum, the CES-D-10, QIDS-SR, and DASS-21-DEP have been shown to be reliable and valid for use in clinical and community samples. However, the CES-D and QIDS-SR have mostly been studied in healthy or outpatient samples, and it is unclear whether any of the three instruments is superior for assessing depression severity and treatment response in an acute, diagnostically heterogeneous, psychiatric population. The present study examined the relative psychometric properties of the CES-D-10, QIDS-SR, and DASS-21-DEP in order to inform selection of an optimal depression measure in such a setting. We evaluated several characteristics of these measures, including internal consistency, discriminative validity, convergent and divergent validity, functionality as a screening tool for depression, and sensitivity to treatment effects. We expected all three measures to show adequate reliability and validity. Convergent and discriminant validity were specifically assessed by Pearson correlations between the depression measures themselves, measures of other closely related constructs (anxiety, worry, stress), and less closely related constructs (overall psychological health, emotion regulation). Data on the additional constructs selected for validity analyses (anxiety, worry, stress, overall psychological health, and emotion regulation) were available to us given that they were collected as part of routine clinical care. We expected

that the CES-D may perform the best as a screening tool given that it was developed for that purpose. However, given the severity of acute populations, we expected that the clinical cut-offs identified in prior research may not be appropriate for this sample. Finally, we expected all three measures to be sensitive to change, but had no a priori hypotheses regarding the superiority of one measure in this domain.

Methods

Participants and Procedures

Participants were 377 patients seeking treatment at a partial hospitalization program utilizing individual and group cognitive behavioral therapy and pharmacotherapy to treat a variety of psychiatric disorders, principally mood, anxiety, personality, and psychotic disorders. The average length of stay in the program is 8.2 (SD=3.2) days. Please see Beard and Björgvinsson (2013) for more details about the treatment and setting. Participants were diagnosed using the Mini International Neuropsychiatric Interview (MINI; Sheehan et al. 1997, 1998). Rates of comorbidity were high, with 55.2 % of the sample meeting criteria for more than one psychiatric diagnosis. See Table 1 for demographic and diagnostic characteristics.

Approval for the study was granted by the hospital's Institutional Review Board, and participants were treated in accordance with the ethical guidelines of the American Psychological Association. All study participants provided written informed consent prior to the study. At admission, participants completed the MINI, a demographics survey, and a battery of self-report measures; the self-report measures were completed again at discharge. Self-report data were collected and managed using REDCap electronic data capture tools hosted at McLean Hospital. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies (Harris et al. 2009).

Measures

Mini International Neuropsychiatric Interview (MINI; Sheehan et al. 1997, 1998). The MINI is a structured interview assessing for Axis I symptoms as outlined by the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV-TR; American Psychiatric Association 2000). The MINI has demonstrated strong reliability with the Structured Clinical Interview for DSM-IV, and inter-rater reliabilities were very high ($\kappa=.89-1.0$; Sheehan et al. 1997). In the present study, the MINI was administered by doctoral practicum students and interns in clinical psychology, who received weekly

Table 1 Demographics and diagnoses ($n=377$)

	<i>M</i> (SD) or <i>N</i> (%)
Age	34.68 (13.01)
Gender	
Male	170 (45.1 %)
Female	207 (54.9 %)
Ethnicity	
American Indian	3 (0.8 %)
Asian	20 (5.3 %)
African American	7 (1.9 %)
Caucasian	315 (83.6 %)
Caribbean Islander	2 (0.5 %)
Latino/a	9 (2.4 %)
Multiracial	9 (2.4 %)
Choose not to respond/Do not know	12 (3.2 %)
Education	
Some high school	4 (1.1 %)
High school graduate	139 (36.9 %)
Some college	39 (10.3 %)
College graduate	83 (22.0 %)
Post-college education	112 (29.7 %)
Employment status	
Employed	176 (46.7 %)
Not employed	201 (53.3 %)
Diagnosis ^a	
Major depressive disorder—currently depressed	216 (57.3 %)
Bipolar disorder—currently depressed	46 (12.2 %)
Major depressive disorder—recurrent (not depressed)	56 (14.9 %)
Bipolar disorder—currently manic or mixed	13 (3.4 %)
Psychotic disorder	22 (5.8 %)
Panic disorder	45 (11.9 %)
Social anxiety disorder	70 (18.6 %)
Obsessive-compulsive disorder	36 (9.5 %)
Posttraumatic stress disorder	37 (9.8 %)
Generalized anxiety disorder	137 (36.3 %)

^a Percentages exceed 100 % due to comorbidity

supervision by a postdoctoral psychology fellow. For the current sample, inter-rater reliability for the MINI and program psychiatrists were within an acceptable range (i.e., .41–1.00; Landis and Koch 1977) for both major depressive disorder ($\kappa=.69$) and bipolar disorder—currently depressed ($\kappa=.75$; Kertz et al. 2012).

Center for Epidemiological Studies Depression Scale-ten Item Version (CES-D-10; Andresen et al. 1994). The CES-D-10 is a brief, widely used self-report instrument used to assess depression symptoms over the past week. Response anchors range temporally from 0=*rarely or none of the time (less than*

1 day) to 3=*most or all of the time (5–7 days)*, with a total possible range of 0–30. The CES-D-10 has demonstrated adequate reliability and validity (Andresen et al. 1994).

Quick Inventory of Depressive Symptoms—Self-Report Version (QIDS-SR; Rush et al. 2003; Trivedi et al. 2004). The QIDS-SR is a 16-item self-report scale that measures the nine criterion symptom domains (sleep, sad mood, appetite/weight, concentration/decision-making, self view, thoughts of death or suicide, general interest, energy level, and restlessness/agitation) that define a DSM-IV-TR major depressive episode over the past week. Six domains are rated by a single item, and the score for the remaining three domains (sleep, appetite/weight, and restlessness/agitation) is derived from the maximum score of two or more questions (e.g., an individual who scores a “2” on an item assessing insomnia and a “0” on an item assessing hypersomnia would receive a total score of “2” for the sleep domain). Each of the nine domains is scored on a Likert scale (representing severity) from zero to three, with a total possible range of zero to 27.

Depression, Anxiety, and Stress Scale-21 Item Version (DASS-21; S. H. Lovibond & Lovibond 1995a, b). The DASS-21 was designed to assess and discriminate between the core symptoms of depression, anxiety, and stress. The 21 items consist of statements about depression (e.g., “I felt that I had nothing to look forward to”), anxiety (e.g., “I was worried about situations in which I might panic and make a fool of myself”), and stress (e.g., “I tended to over-react to situations”). Participants rate how true each statement has been true of them in the past week on a four-point Likert scale, from 0 (*did not apply to me*) to 3 (*applied to me very much or most of the time*).

The 7-item Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al. 2006) is a self-report questionnaire that was developed in a primary care patient sample specifically to increase recognition of GAD (Spitzer et al. 2006). Participants are asked how often in the past two weeks they have been bothered by each of the main characteristics of GAD (e.g., trouble relaxing). Participants respond according to a four-point Likert scale, from 0 (*not at all*), to 3 (*nearly every day*). The total possible range of scores is from zero to 21, where higher scores indicate greater severity of anxiety symptoms. The GAD-7 has demonstrated good reliability and construct validity, as evidenced by its associations with depression, self-esteem, life satisfaction, and resilience (Kroenke et al. 2007; Löwe et al. 2008; Spitzer et al. 2006). In the present sample, internal consistency of the GAD-7 was high ($\alpha=0.90$).

Penn State Worry Questionnaire-Abbreviated (PSWQ-A; Hopko et al. 2003). The PSWQ-A is a well-validated, single

factor, eight-item measure designed to assess worry severity. Derived from Meyer et al. (1990) original 16-item instrument, items on the PSWQ-A consist of statements about worry (e.g., “Many situations make me worry”) that participants rate on a five-point Likert type rating scale ranging from 1 (*not at all typical of me*) to 5 (*very typical of me*) with no specified time frame. Total scores range from 8 to 40, with higher scores indicating higher levels of worry. In the present sample, internal consistency of the PSWQ-A was high ($\alpha=0.94$).

Schwartz Outcome Scale (SOS; Blais et al. 1999). The SOS is a well-validated and reliable, single factor, ten-item measure designed to examine the broad domain of psychological health in a variety of settings (Young et al. 2003). Each item assesses for psychological well-being (e.g., “My life is according to my expectations”). Participants rate items on a seven-point Likert scale from 0 (*never*) to 6 (*all or nearly all of the time*) for the past week. Total scores range from 0 to 60, with higher scores indicating better psychological health. In the present sample, internal consistency of the SOS was high ($\alpha=0.93$).

Emotion Regulation Questionnaire (ERQ; Gross and John 2003). The ERQ is a 10-item self report inventory assessing individual competencies on emotion regulation strategies. Each item consists of a statement that participants rate on a seven-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*) according to how they have generally been feeling. The ERQ is divided into the two subscales of reappraisal (e.g., “When I want to feel more positive emotion (such as joy or amusement), I change what I’m thinking about”) and suppression (e.g., “I keep my emotions to myself”). Total ERQ scores were used in the present study. In the current sample, Cronbach’s alpha for ERQ total scores was $\alpha=.77$.

Analytic Strategy

Of the 17,719 item responses possible in the present study, 127 item responses had missing data (0.7 %). Following common guidelines in data imputation, we did not impute these data using multiple imputation or hot deck imputation as these advanced techniques are more suitable for datasets with higher percentages of missing values (>5 or >10 %; Tabachnick and Fidell 2004; Myers 2011). Instead we conducted all analyses on available data only without imputing missing values.

We first examined any demographic differences on mean scores for each measure. Internal consistency was assessed by coefficient alphas, item-total correlations and inter-item correlations for the three depression measures. We identified an

acceptable alpha value as above .70. We required strong correlations ($r>.50$ among the depression measures) for a measure to show acceptable convergent validity. We required correlations to be weak ($r<.30$) as acceptable support for discriminant validity. We expected correlations between the depression measures and measures of anxiety, worry, and stress to be moderate (r ’s between .30 and .50). Divergent validity was assessed by comparing scores between patients with and without a current Major Depressive Episode (MDE). A receiving operator characteristic (ROC) curve was estimated to examine whether the three depression measures could distinguish between individuals with and without a MDE across the range of cut-off scores. Finally, sensitivity to treatment was examined by completer and intent-to treat (ITT) analyses of pre-post mean differences using ANOVA, as well as comparing the number of patients experiencing a reliable change on each measure.

Results

Descriptive Statistics

Table 2 presents means and standard deviations for the three depression measures, as well as for measures of convergent and discriminant validity. Females scored significantly higher on all three depression measures at baseline ($F(1, 375)=8.88, p<0.01$, partial $\eta^2=0.02$; $F(1, 375)=6.68, p<0.05$, partial $\eta^2=0.02$; $F(1, 375)=4.07, p<0.05$, partial $\eta^2=0.01$, for the QIDS-SR, CES-D-10, and DASS-21-DEP respectively). The CES-D-10, QIDS-SR and DASS-21-DEP were also significantly positively associated with age (r ’s=.12–.16, p ’s<.05). Each measure significantly differed on marital status, such that individuals who were separated, divorced, or widowed reported more depressive symptoms (CES-D-10 $M=18.37, SD=7.26$; QIDS-SR $M=14.87, SD=6.59$; and DASS-21-DEP $M=23.73, SD=12.24$) compared to individuals who were never married (CES-D-10 $M=14.88, SD=7.82$; QIDS-SR $M=12.22, SD=6.23$; DASS-21-DEP $M=17.21, SD=13.11$; $F(1, 375)=5.68, p<0.05$, partial $\eta^2=0.02$; $F(1, 375)=4.42, p<0.05$, partial $\eta^2=0.01$; $F(1, 375)=6.78, p<0.05$, partial $\eta^2=0.02$, for the QIDS-SR, CES-D-10, and DASS-21-DEP respectively). Participants with a higher level of education reported more depressive symptoms on the CES-D-10 ($F(1, 375)=3.94, p<0.05$, partial $\eta^2=0.01$) and the DASS-21-DEP ($F(1, 375)=4.80, p<0.05$, partial $\eta^2=0.01$), but not on the QIDS-SR ($F(1, 375)=1.22, p=0.05$, partial $\eta^2=0.00$). No significant relationships emerged between any of the depression measures and race/ethnicity (White Non-Latino/a vs. African American/Latino/a/other) or employment status.

Table 2 Means and standard deviations for clinical measures

	Pre-treatment								
	Total sample			Current MDE			No current MDE		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
CES-D-10	377	15.57	7.85	277	17.66	7.41	100	9.78	5.92
QIDS-SR	377	12.84	6.23	277	14.38	5.87	100	8.58	5.13
DASS-21-DEP	377	18.64	13.14	277	22.19	12.52	100	8.80	9.25
DASS-21-ANX	371	9.58	7.84	272	10.74	8.07	99	6.42	6.23
DASS-21-STR	371	14.53	9.67	272	15.90	9.93	99	10.77	7.83
GAD-7	368	9.45	6.02	271	10.49	6.13	97	6.55	4.65
PSWQ-A	365	28.57	8.72	267	30.16	8.28	98	24.24	8.46
SOS	364	31.62	14.96	267	27.88	12.47	97	41.92	16.40
ERQ	362	36.25	9.55	265	34.03	9.12	97	39.57	9.95
CES-D-10	377	11.73	7.37	277	13.27	7.53	100	7.45	4.79
QIDS-SR	377	9.38	6.07	277	10.64	6.09	100	5.89	4.42
DASS-21-DEP	377	13.59	12.37	277	16.40	12.63	100	5.82	7.33
DASS-21-ANX	371	7.10	6.96	272	8.18	7.22	99	4.16	5.15
DASS-21-STR	371	11.36	9.19	272	12.69	9.51	99	7.72	7.10
GAD-7	368	6.91	5.54	267	7.84	5.73	98	4.34	4.03
PSWQ-A	365	25.98	9.07	271	27.72	8.56	97	21.18	8.74
SOS	364	38.55	15.61	267	35.09	14.57	97	48.03	14.44
ERQ	362	39.42	8.13	265	38.32	7.89	97	42.44	8.05

MDE Major Depressive Episode, *CES-D-10* Center for Epidemiological Studies Depression Scale-ten item version, *QIDS-SR* Quick Inventory of Depressive Symptomatology-Self-report version, *DASS-21-DEP* Depression Anxiety and Stress Scales-21 item version, depression subscale, *DASS-21-ANX* Depression Anxiety and Stress Scales-21 item version, anxiety subscale, *DASS-21-STR* Depression Anxiety and Stress Scales-21 item version, stress subscale, *GAD-7* Generalized Anxiety Disorder Scale, 7 item version, *PSWQ-A* Penn State Worry Questionnaire-Abbreviated, *SOS* Schwartz Outcome Scale, *ERQ* Emotion Regulation Questionnaire

Reliability

The total scores for the three depression measures demonstrated high internal consistency (Cronbach's α 's = .90, .81, and .94 for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively). Item-total correlations for the CES-D-10 ranged from moderate ("I felt hopeful about the future"; $r = .58$, $p < .001$) to high ("I felt depressed"; $r = .87$, $p < .001$). Item-total correlations for the QIDS-SR ranged from low ("Waking up too early"; $r = .24$, $p < .001$) to high ("Energy Level"; $r = .80$, $p < .001$). Item-total correlations for the DASS-21-DEP were in the high range ("I found it difficult to work up the initiative to do things"; $r = .82$, $p < .001$, to "I felt that I had nothing to look forward to"; $r = .91$, $p < .001$). Inter-item correlations for the three depression measures are provided in Appendix 1. Average inter-item correlations for depression measures were 0.48 (range = 0.19–0.75), 0.22 (range = -0.26–0.69), and 0.40 (range = 0.15–0.77) for the CES-D-10, QIDS-SR, and DASS-21-DEP respectively. Although we expected total scores to change in response to treatment, we examined pre-post correlations as an indicator of test-retest reliability. Pre-post

correlations were 0.60, 0.69, and 0.66 for the CES-D-10, QIDS-SR, and DASS-21-DEP.

Convergent Validity

Convergent validity was assessed by examining the correlations between the three depression measures (see Table 3). All three depression measures showed strong correlations with one another (r 's range from .83 to .86).

Discriminant and Divergent Validity

We first examined measures of clinically related constructs, i.e., anxiety, as we expected significant, although relatively weaker, correlations between these measures and the depression measures. All three depression measures showed the same pattern of correlations, with strong, positive correlations with the DASS-21-ANX, DASS-21-STR, and GAD-7, and moderate correlations with the PSWQ-A. Using Meng et al. (1992) formula for comparing between correlated correlation coefficients, we found that all correlations between the depression measures on one hand and related measures of

Table 3 Pearson correlations between clinical measures

	1	2	3	4	5	6	7	8	9
1. CES-D-10	–								
2. QIDS	.86***	–							
3. DASS-21-DEP	.86***	.83***	–						
4. PSWQ	.37***	.37***	.34***	–					
5. SOS	–.28***	–.21***	–.32***	–.51***	–				
6. DASS-21-ANX	.63***	.64***	.58***	.34***	.07	–			
7. DASS-21-STR	.72***	.69***	.67***	.34***	–.11*	.72***	–		
8. GAD-7	.75***	.72***	.64***	.45***	–.09	.76***	.80***	–	
9. ERQ	–.27***	–.27***	–.30***	.32***	.42***	–.12*	–.19***	.21***	–

* $p < .05$, ** $p < .01$, *** $p < .001$

CES-D-10 Center for Epidemiological Studies Depression Scale-ten item version, *QIDS* Quick Inventory of Depressive Symptomatology, *DASS-21-DEP* Depression Anxiety Stress Scales-21 item version, depression subscale, *PSWQ-A* Penn State Worry Questionnaire-Abbreviated, *SOS* Schwartz Outcome Scale, *DASS-21-ANX* Depression Anxiety Stress Scales-21 item version, anxiety subscale, *DASS-21-STR* Depression Anxiety Stress Scales-21 item version, stress subscale, *GAD-7* Generalized Anxiety Disorder Scale, 7 item version, *ERQ* Emotion Regulation Questionnaire

anxiety, stress and worry on the other were significantly smaller than the correlations between the depression measures themselves (all $ps < 0.001$).

To further examine discriminant validity, we selected measures of constructs that should be relatively less related to depression, i.e., psychological well-being (SOS), and emotion regulation skills (ERQ). Results are presented in Table 3. The CES-D-10, QIDS-SR, and DASS-21-DEP demonstrated small, negative associations with the SOS ($r = -.21$ – $-.32$, p 's $< .001$) and ERQ ($r = -.27$ – $-.30$, p 's $< .001$) scores. Using Meng et al. (1992) formula for comparing between correlated correlation coefficients, we found that all correlations between the depression measures on one hand and SOS and ERQ on the other were significantly smaller than the correlations between the depression measures themselves (all $ps < 0.001$). The correlations between vectors of correlations of the three depression measures on one hand and all other measures was nearly perfect (all $rs > 0.99$). This indicates that the patterns of correlations of the three measures are nearly identical.

We assessed divergent validity by comparing scores for individuals with and without a current MDE as assessed by the MINI. Significant differences emerged on all three depression measures. Specifically, individuals with a current MDE scored significantly higher on the CES-D-10, the QIDS-SR, and the DASS-21-DEP at pre-treatment ($F(1, 375) = 76.30$, $p < 0.001$, partial $\eta^2 = 0.17$; $F(1, 375) = 91.91$, $p < 0.001$, partial $\eta^2 = 0.20$; $F(1, 375) = 95.52$, $p < 0.001$, partial $\eta^2 = 0.20$).

Cut-scores, Sensitivity, and Specificity

A receiving operator characteristic (ROC) curve was estimated to examine whether the three depression measures could distinguish between individuals with and without a MDE

across the range of cut-off scores. A ROC curve indicated that the areas under the curve were .79 (95 % CI=.74–.84), .77 (95 % CI=.72–.83), and .79 (95 % CI=.75–.84) for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively. In our acute psychiatric sample, the optimal cut-off score for the CES-D-10 was 16, with adequate sensitivity (.67) and good specificity (.81) for identifying individuals with a current MDE. The optimal cut-off score for the QIDS-SR was 12, with adequate sensitivity (.74) and adequate specificity (.77). The optimal cut-off score for the DASS-21-DEP was 19, with adequate sensitivity (.60) and good specificity (.88). Table 4 presents sensitivity and specificity scores for various cut-off points for the three depression measures.

Sensitivity to Treatment

We examined sensitivity to treatment by comparing pre- and post-treatment scores for individuals with and without a MDE on all three depression measures (see Table 2) using a mixed ANOVA. This analysis was based on completer data with a sample of 270 individuals at post-treatment. The analyses yielded a significant pre-post effect for all three measures ($F(1,268) = 162.86$, $p < .001$; $F(1,268) = 272.94$, $p < .001$; and $F(1,268) = 130.32$, $p < .001$, for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively) such that all measures showed a decrease in symptoms from pre- to post-treatment. In addition, a significant effect of MDE was found ($F(1,268) = 851.73$, $p < .001$; $F(1,268) = 851.66$, $p < .001$; and $F(1,268) = 373.84$, $p < .001$, for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively) such that individuals with a MDE scored significantly higher on all three measures. Finally, a significant treatment times MDE interaction was found ($F(1,268) = 11.39$, $p < .01$; $F(1,268) = 6.26$, $p < .05$; and $F(1,268) = 12.17$,

Table 4 Sensitivity and specificity of cut-off scores for identifying a current major depressive episode

Cut-off score	Sensitivity	Specificity
CES-D-10		
9	.86	.47
10	.85	.55
11	.83	.58
12	.80	.62
13	.75	.68
14	.73	.71
15	.70	.76
16	.67	.81
17	.62	.84
18	.58	.90
19	.53	.91
QIDS-SR		
9	.83	.53
10	.81	.60
11	.78	.70
12	.74	.77
13	.70	.79
14	.62	.81
15	.54	.84
16	.47	.89
17	.38	.92
18	.34	.95
19	.26	.96
DASS-21-DEP		
9	.83	.61
11	.79	.65
13	.75	.71
15	.70	.77
17	.63	.84
19	.60	.88
21	.56	.89
23	.50	.90
25	.45	.91
27	.39	.94

CES-D-10 Center for Epidemiological Studies Depression Scale-ten item version, *QIDS-SR* Quick Inventory of Depressive Symptomatology-Self-report version, *DASS-21-DEP* Depression Anxiety and Stress Scales-21 item version, depression subscale

$p < .01$, for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively) such that individuals with a MDE improved more during treatment compared to individuals without a MDE. In order to obtain a measure of effect size for sensitivity to treatment, we computed Cohen's d for pre-post differences. Effect sizes were medium for all three measures (Cohen's $d = .68, .76, \text{ and } .60$ for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively).

We also conducted an ITT analysis using the Last-Observation-Carried-Forward (LOCF; Shao and Zhong 2003) technique and including all participants ($n = 377$). Similarly, analyses yielded a significant pre-post effect for

all three measures ($F(1, 376) = 149.00, p < 0.001$; $F(1, 376) = 224.29, p < 0.001$; $F(1, 376) = 121.82, p < 0.001$; for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively). Effect sizes for the ITT analyses were medium for all three measures (Cohen's $d = .50, .56, \text{ and } .60$ for the CES-D-10, QIDS-SR, and DASS-21-DEP, respectively).

Finally, we used Jacobson and Traux (1991) criteria to examine reliable change in depressive measures. The average reliable change index was 1.10, 1.27, and 0.88 for the CES-D-10, QIDS-SR, and DASS-21-DEP respectively. The cutoff for significant reliable change is 1.96. Percentages of individuals classified as having reliable change were 26.7, 25.4, and 19.1 % for the CES-D-10, QIDS-SR, and DASS-21-DEP respectively. We also found significant differences in the average reliable change index among individuals with and without a MDE such that individuals with a MDE had larger values (i.e., improved more) than individuals without a MDE (all $ps < 0.01$).

Discussion

The aims of the present study were to compare the psychometric properties of three brief self-report measures of depression, and to evaluate their relative utility in an acute psychiatric partial hospital setting. Although the measures were developed for different purposes, results indicated that all three have highly acceptable psychometric properties in this population and setting. As one might expect, mean total scores in the current partial hospital sample were generally higher than those reported in prior studies. It is difficult to compare the CESD-10 means in the present sample to typical depressed outpatient samples because most studies utilizing the ten-item version of the CES-D have included specific medical populations, older community populations, or adolescents. However, our mean ($M = 15.57$) was higher than those reported in other studies using this measure (e.g., older community sample $M = 4.7$; Andresen et al. 1994). Similarly, the QIDS mean in the current partial hospital sample ($M = 12.84$) was higher than reported means for depressed outpatients ($M = 7.7$; Rush et al. 2006a, b). However, the DASS mean in the current partial hospital sample ($M = 18.74$) was lower than that reported for inpatient samples (e.g., $M = 29.01$; Ng et al. 2007) and outpatient depressed samples (29.96; Antony et al. 1998).

With respect to internal consistency, performance was excellent in the DASS-21-DEP and the CES-D-10, and good in the QIDS-SR. The internal consistency of the QIDS-SR in the present sample was very similar to that previously reported in depressed outpatients (Rush et al. 2006a; Rush et al. 2003; Trivedi et al. 2004). For the CES-D-10, internal consistency

was higher than that reported in previous studies (Boey 1999; Bradley et al. 2010; Lee and Chokkanathan 2008). The internal consistency of the DASS-21-DEP was equal to or higher than that reported in other clinical and nonclinical samples (e.g., Antony et al. 1998; Gloster et al. 2008; Henry and Crawford 2005; Norton 2007). The DASS-21-DEP also performed best at the item-level, with all item-total correlations in the high range. This may reflect the fact that the DASS-21-DEP focuses on cognitive and emotional aspects of depressive symptomatology, whereas the CES-D-10 and QIDS-SR also measure neurovegetative or behavioral features of depression. Of the three depression measures, the QIDS-SR is most comprehensive in its assessment of multiple domains of depressive symptoms (cognitive, emotional, and behavioral/neurovegetative). Lower average inter-item correlations are therefore to be expected in the QIDS-SR, given that it assesses a more heterogeneous set of facets. Overall, the three measures demonstrated good to excellent internal consistency, and performed comparably relative to internal consistencies reported in a variety of other samples.

All three measures also demonstrated strong convergent and discriminant validity in this sample. All were highly positively correlated with each other, and were also positively correlated with measures of anxiety and stress. The depression measures demonstrated positive correlations with the construct of worry, but the strength of association was weaker than with more general anxiety and stress constructs. This finding was expected, given that the PSWQ assesses the specific construct of worry (a purely cognitive process), as opposed to more global measures of anxiety and stress. Supporting the discriminant validity of the depression measures, all correlations between the depression measures on one hand and related measures of anxiety, stress and worry on the other, were significantly smaller than the correlations between the depression measures themselves. In further tests of discriminant validity, the depression measures demonstrated relatively weaker negative correlations with the constructs of psychological well-being and emotion regulation skills. The CES-D-10, QIDS-SR, and DASS-21-DEP also demonstrated divergent validity, with individuals currently experiencing a MDE scoring higher than those who were not currently depressed. These findings indicate that the three measures are sensitive to a diagnosis of major depression even in a highly comorbid, acutely symptomatic sample. Thus, consistent with our hypotheses, the depression measures were positively associated with each other and with measures assumed to represent similar constructs. We examined diagnostic screening utility with respect to the prediction of the presence of a major depressive episode. These analyses indicated that the three measures predict the presence of a major depressive episode equally well, and that their diagnostic utilities are fair to good. Regarding the CES-D-10, previously suggested cut-off

scores of 8 and 10 (Andresen et al. 1994; Boey 1999) or 12 to 13 (Cheng and Chan 2005) performed poorly in the present sample; a cut-off score of 16 optimized sensitivity and specificity. For the QIDS-SR, the AUC and optimal cut-off scores were higher than those reported in a sample of elderly adults (Doraiswamy et al. 2010), and similar to those reported in a primary care sample (Lamoureux et al. 2010; Lee and Chokkanathan 2008). Regarding the DASS-21-DEP, the AUC in the present sample was similar to that found in a sample of older adults in a primary care setting (Gloster et al. 2008), but the optimal cut-off score was higher. Thus, the three measures demonstrated comparable diagnostic utility in comparison to previous studies, although optimal cut-off scores were generally higher in the present sample. Clinicians and researchers should use caution when utilizing previously recommended cutoffs for the CES-D-10, QIDS-SR, and DASS-21-DEP within acute psychiatric settings. In such settings, higher cut-offs may be needed in order to maximize sensitivity and specificity for detecting a depression diagnosis.

All three measures were sensitive to changes in depressive symptomatology across the course of treatment as evidenced by medium pre-post changes and a similar percentage of patients experiencing a reliable change. The QIDS-SR was the most sensitive to treatment effects, which is consistent with extant literature and with the purpose of its original development (Rush et al. 2006a; Rush et al. 2006b; Rush et al. 2003; 2005; Trivedi et al. 2004). Given that the DASS-21-DEP and the CES-D-10 have been less well studied in clinical samples, fewer psychometric evaluations have examined the treatment sensitivity of these measures. One study in particular found the DASS-21-DEP to be sensitive to change in a private inpatient setting (Ng et al. 2007). Results suggest that the QIDS-SR, DASS-21-DEP, and CES-D-10 are all capable of reflecting change in clinical status across treatment in an acute psychiatric population.

The results of this study indicate that any of the three depression measures may satisfactorily assess depressive symptoms in an acute psychiatric population. Thus, the selection of a specific assessment tool should be guided by the identified purpose of the assessment (see Joiner et al. 2005). In a partial hospital setting, the QIDS-SR may confer some advantages over the other two measures. Specifically, the QIDS-SR may be preferable for the purpose of diagnostic screening, as the items correspond directly to DSM criteria. The QIDS-SR was also most sensitive to change during the course of treatment, and may therefore be preferable as a treatment outcome measure. Also, of the three depression measures examined in the present study, only the QIDS-SR explicitly measures suicidality. In reviewing the literature on evidence-based assessment of depression, Joiner et al. (2005) noted that the assessment of suicidality is critical in

differentiating major depression from distress, demoralization, and sad mood.

In front-line clinical settings, it may be necessary to utilize the briefest, most feasible instrument that offers sufficient reliability and validity (Barlow 2005). The CES-D-10 meets this criterion in that it is briefer than the QIDS-SR and the full DASS-21. The DASS-21-DEP contains only seven items, but it has not been validated for use as a stand-alone measure. The DASS-21-DEP may be preferable if there is a concurrent interest in anxiety and the other DASS-21 subscales can be administered, given that the anxiety and stress scales have performed well in clinical samples (e.g., Antony et al. 1998; Clara et al. 2001; Ng et al. 2007). However, it is important to note that the DASS-21-DEP fails to assess neurovegetative or behavioral aspects of depression.

Several limitations should be noted when interpreting results of the present study. First, given the time constraints and concern about patient burden in our acute psychiatric setting, we could only compare three measures in this study. Second, all measures used to examine convergent and discriminant validity were based on subjective report. The inclusion of behavioral or neurobiological measures of depression would have provided a more rigorous test of convergent validity. Related to this is our focus on comparing self-report measures. Although brief, self-report measures are certainly the most feasible to administer in real-world, acute settings, ideally clinicians would also administer clinician-rated scales as both types of assessment appear to provide unique information (Uher et al. 2012). Third, although the sample was diagnostically heterogeneous, it was relatively homogeneous in terms of ethnicity. Future work should compare the psychometric properties of these brief depression measures in a sample with

more ethnic and racial diversity. Fourth, evidence regarding dimensionality of the self report measures, as well as measurement invariance among clinical and non-clinical samples, males and females, and pre-treatment as well as post-treatment measurements is not provided in the current study. Comparing sum scores between these measures may have weaknesses if these assumptions are not met. Future studies can use item level psychometric analyses which may uncover differences among the instrument item sets that may be obscured when using aggregated sum score and correlation based summaries. Finally, in the context of this naturalistic treatment study, we lacked a control group of participants against which to compare sensitivity to treatment effects.

Despite these limitations, the present study is the first to evaluate and compare the psychometric properties of these three brief depression measures in a naturalistic, diagnostically heterogeneous, psychiatrically acute partial hospitalization sample. Although the three self-report measures were developed for different purposes, all three performed well in this patient population. Thus, selection of a particular measure should be embedded within the purpose of the assessment. Future research should examine whether the three measures perform differently across the spectrum of symptom severity. Moreover, it will be important to determine whether these measures confer incremental utility over clinician judgment alone, and are therefore able to improve treatment outcomes.

Conflict of Interest There are no conflicts of interest to disclose.

Experiment Participants The study was approved by the Institutional Review Board and all participants provided written informed consent.

Appendix

Table 5 Inter-Item Correlations of the CES-D

Item	1	2	3	4	5	6	7	8	9	10
1	1									
2	.524***	1								
3	.557***	.617***	1							
4	.507***	.647***	.754***	1						
5	.245***	.305***	.468***	.419***	1					
6	.429***	.465***	.563***	.566***	.264***	1				
7	.369***	.463***	.479***	.485***	.189***	.433***	1			
8	.400***	.346***	.569***	.494***	.599***	.288***	.262***	1		
9	.453***	.475***	.620***	.570***	.407***	.434***	.379***	.451***	1	
10	.488***	.604***	.684***	.681***	.431***	.454***	.454***	.466***	.640***	1

*** $p < .001$

Table 6 Inter-Item Correlations of the QIDS-SR

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1															
2	.383***	1														
3	.182***	.337***	1													
4	.049	.064	-.163**	1												
5	.258***	.289***	.125*	.251***	1											
6	.188***	.237***	.143**	.031	.290***	1										
7	.097	.153**	-.010	.153**	.247***	-.239***	1									
8	-.023	.093	.137*	-.066	.146**	.524***	-.147**	1								
9	.106*	.160**	.028	.068	.098	-.255***	.552***	-.261***	1							
10	.294***	.364***	.222***	.206**	.512***	.270***	.262***	.120*	.207***	1						
11	.147**	.194***	.050	.263***	.543***	.172**	.229***	.055	.095	.416***	1					
12	.130*	.190***	.056	.193***	.473***	.145**	.163**	.078	.039	.241***	.389***	1				
13	.253***	.286***	.065	.315***	.614***	.357***	.195***	.150**	.140***	.579***	.509***	.298***	1			
14	.353***	.328***	.149**	.395***	.592***	.317***	.277***	.106*	.178**	.602***	.470***	.276***	.692***	1		
15	.214***	.274***	.174**	.197***	.423***	.257***	.229***	.181**	.199***	.549***	.305***	.269***	.426***	.490***	1	
16	.143**	.174**	.134**	.072	.172**	.111*	.191***	.165**	.179**	.350***	.194***	.152**	.236***	.199***	.246***	1

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7 Inter-Item Correlations of the DASS-21

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21		
1	1																						
2	.298***	1																					
3	.383***	.313***	1																				
4	.278***	.229***	.230***	1																			
5	.334***	.279***	.698***	.271***	1																		
6	.370***	.248***	.499***	.311***	.513***	1																	
7	.234***	.151***	.257***	.260***	.321***	.333***	1																
8	.431***	.208***	.362***	.251***	.389***	.492***	.415***	1															
9	.358***	.283***	.360***	.331***	.387***	.402***	.293***	.536***	1														
10	.304***	.222***	.746***	.226***	.686***	.489***	.290***	.305***	.381***	1													
11	.365***	.219***	.521***	.308***	.463***	.655***	.283***	.451***	.401***	.492***	1												
12	.571***	.287***	.577***	.261***	.484***	.482***	.270***	.505***	.417***	.478***	.536***	1											
13	.321***	.265***	.740***	.250***	.736***	.506***	.267***	.368***	.407***	.766***	.522***	.504***	1										
14	.259***	.224***	.391***	.240***	.353***	.498***	.268***	.343***	.292***	.398***	.530***	.385***	.383***	1									
15	.409***	.291***	.454***	.365***	.449***	.530***	.312***	.520***	.544***	.486***	.571***	.529***	.495***	.475***	1								
16	.292***	.198***	.762***	.258***	.731***	.437***	.232***	.253***	.325***	.733***	.459***	.522***	.774***	.328***	.457***	1							
17	.317***	.204***	.637***	.206***	.573***	.490***	.297***	.311***	.367***	.747***	.426***	.480***	.682***	.315***	.438***	.649***	1						
18	.340***	.235***	.506***	.210***	.406***	.550***	.252***	.286***	.283***	.459***	.526***	.424***	.483***	.377***	.437***	.416***	.491***	1					
19	.300***	.216***	.274***	.463***	.278***	.324***	.271***	.308***	.371***	.292***	.325***	.318***	.304***	.318***	.419***	.237***	.326***	.303***	1				
20	.398***	.226***	.448***	.268***	.398***	.380***	.224***	.421***	.482***	.401***	.373***	.477***	.418***	.288***	.553***	.405***	.422***	.336***	.441***	1			
21	.267***	.200***	.673***	.203***	.561***	.408***	.251***	.270***	.298***	.764***	.439***	.424***	.676***	.400***	.457***	.687***	.711***	.418***	.250***	.372***	1		

p < .01, *** *p* < .001

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC.
- Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: evaluation of a short form of the CES-D. *American Journal of Preventive Medicine, 10*(2), 77–84.
- Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181.
- Barlow, D. H. (2005). What's new about evidence-based assessment? *Psychological Assessment, 17*(3), 308–311.
- Beard, C., & Björgvinsson, T. (2013). Commentary on psychological vulnerability: an integrative approach. *Journal of Integrative Psychotherapy, 23*(3), 281–283.
- Björgvinsson, T., Kertz, S. J., Bigda-Peyton, J. S., McCoy, K. L., & Aderka, I. M. (2013). Psychometric properties of the CES-D-10 in a naturalistic psychiatric sample. *Assessment, 20*(4), 429–436.
- Blais, M. A., Lenderking, W. R., Baer, L., deLorell, A., Peets, K., Leahy, L., & Burns, C. (1999). Development and initial validation of a brief mental health outcome measure. *Journal of Personality Assessment, 73*(3), 359–373.
- Boey, K. W. (1999). Cross-validation of a short form of the CES-D in Chinese elderly. *International Journal of Geriatric Psychiatry, 14*(8), 608–617.
- Bradley, K. L., Bagnell, A. L., & Brannen, C. L. (2010). Factorial validity of the center for epidemiological studies depression 10 in adolescents. *Issues in Mental Health Nursing, 31*(6), 408–412.
- Cheng, S.-T., & Chan, A. C. M. (2005). The center for epidemiologic studies depression scale in older Chinese: thresholds for long and short forms. *International Journal of Geriatric Psychiatry, 20*(5), 465–470.
- Cheng, S.-T., Chan, A. C. M., & Fung, H. H. (2006). Factorial structure of a short version of the center for epidemiologic studies depression scale. *International Journal of Geriatric Psychiatry, 21*(4), 333–336.
- Cheung, Y. B., Liu, K. Y., & Yip, P. S. F. (2007). Performance of the CES-D and its short forms in screening suicidality and hopelessness in the community. *Suicide and Life-Threatening Behavior, 37*(1), 79–88.
- Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory factor analysis of the depression-anxiety-stress scales in depressed and anxious patients. *Journal of Psychopathology and Behavioral Assessment, 23*(1), 61–67.
- Doraiswamy, P. M., Bernstein, I. H., Rush, A. J., Kyutoku, Y., Carmody, T. J., Macleod, L., & Trivedi, M. H. (2010). Diagnostic utility of the quick inventory of depressive symptomatology (QIDS-C16 and QIDS-SR16) in the elderly. *Acta Psychiatrica Scandinavica, 122*(3), 226–234.
- Duffy, F. F., Chung, H., Trivedi, M., et al. (2008). Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatric Services, 59*, 1148–1154.
- Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Psychiatrists in the UK do not use outcome measures. National Survey. *British Journal of Psychiatry, 180*, 101–103.
- Gloster, A. T., Rhoades, H. M., Novy, D., Klotsche, J., Senior, A., Kunik, M., & Stanley, M. A. (2008). Psychometric properties of the depression anxiety and stress scale-21 in older primary care patients. *Journal of Affective Disorders, 110*(3), 248–259.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348–362.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, J., & Conde, D. (2009). Research electronic data capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381.
- Henry, J. D., & Crawford, J. R. (2005). The short-form version of the depression anxiety stress scales (DASS-21): construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(2), 227–239.
- Hopko, D. R., Reas, D. L., Beck, J. G., Stanley, M. A., Wetherell, J. L., Novy, D. M., & Averill, P. M. (2003). Assessing worry in older adults: confirmatory factor analysis of the penn state worry questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment, 15*(2), 173–183.
- Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment, 17*(3), 251–255.
- Jacobson, N. S., & Traux, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- 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*(3), 267–277.
- Kertz, S. J., Bigda-Peyton, J. S., Rosmarin, D. H., & Björgvinsson, T. (2012). Worry across diagnostic presentations: prevalence, severity, and associated symptoms in a partial hospital setting. *Journal of Anxiety Disorders, 26*, 126–133.
- Kroenke, K., Spitzer, R. L., Williams, J. B., Monohan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*, 317–325.
- Lamoureux, B. E., Linardatos, E., Fresco, D. M., Bartko, D., Logue, E., & Milo, L. (2010). Using the QIDS-SR16 to identify major depressive disorder in primary care medical patients. *Behavior Therapy, 41*(3), 423–431.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
- Lee, A. E. Y., & Chokkanathan, S. (2008). Factor structure of the 10-item CES-D scale among community dwelling older adults in Singapore. *International Journal of Geriatric Psychiatry, 23*(6), 592–597.
- Lovibond, P. F., & Lovibond, S. H. (1995a). The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343.
- Lovibond, S. H., & Lovibond, P. F. (1995b). *Manual for the depression anxiety stress scales* (2nd ed.). Sydney: Psychology Foundation.
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274.
- Meng, X., Rosenthal, R., & Rubin D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 112*, 172–175.
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28*(6), 487–495.
- Myers. (2011). Goodbye, listwise deletion: presenting Hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures, 5*, 297–310. doi:10.1080/19312458.2011.624490.
- Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., & Berk, M. (2007). The validity of the 21-item version of the depression anxiety stress scales as a routine clinical outcome measure. *Acta Neuropsychiatrica, 19*(5), 304–310.
- Norton, P. J. (2007). Depression anxiety and stress scales (DASS-21): psychometric analysis across four racial groups. *Anxiety, Stress & Coping: An International Journal, 20*(3), 253–265.
- Patrick, J., Dyck, M., & Bramston, P. (2010). Depression anxiety stress scale: is it valid for children and adolescents? *Journal of Clinical Psychology, 66*(9), 996–1007.

- Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401.
- Ronk, F. R., Korman, J. R., Hooke, G. R., & Page, A. C. (2014). Assessing clinical significance of treatment outcomes using the DASS-21. *Psychological Assessment*. doi:10.1037/a0033100.
- Rush, A. J., Gullion, C. M., Basco, M. R., & Jarrett, R. B. (1996). The inventory of depressive symptomatology (IDS): psychometric properties. *Psychological Medicine*, *26*(3), 477–486.
- Rush, A. J., Carmody, T., & Reimitz, P.-E. (2000). The inventory of depressive symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *Journal of Practical Psychiatry and Behavioral Health*, *9*, 45–59.
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., & Keller, M. B. (2003). The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583.
- Rush, A. J., Trivedi, M. H., Carmody, T. J., Ibrahim, H. M., Markowitz, J. C., Keitner, G. I., & Keller, M. B. (2005). Self-reported depressive symptom measures: sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients. *Neuropsychopharmacology*, *30*(2), 405–416.
- Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., & Fava, M. (2006a). An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry*, *59*(6), 493–501.
- Rush, A. J., Carmody, T. J., Ibrahim, H. M., Trivedi, M. H., Biggs, M. M., Shores-Wilson, K., & Kashner, T. M. (2006b). Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services*, *57*(6), 829–837.
- Shao, J., & Zhong, B. (2003). Last observation carry-forward and last observation analysis. *Statistics in Medicine*, *22*, 2429–2441.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., & Dunbar, G. C. (1997). The validity of the mini international neuropsychiatric interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*(5), 232–241.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., & Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(Suppl 20), 22–33.
- Slade, M., McCrone, P., Kuipers, E., et al. (2006). Use of standardised outcome measures in adult mental health services: randomised controlled trial. *British Journal of Psychiatry*, *189*, 330–336.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.
- Tabachnick, B. G., & Fidell, L. S. (2004). *Using multivariate statistics* (5th ed.). Boston: Pearson Publications.
- Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., & Kashner, T. M. (2004). The inventory of depressive symptomatology, clinician rating (IDS-C) and self-report (IDS-SR), and the quick inventory of depressive symptomatology, clinician rating (QIDS-C) and self-report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychological Medicine*, *34*(1), 73–82.
- Uher, R., Perlis, R. H., Placentino, A., Dernovšek, M. Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., & Farmer, A. (2012). Self-report and clinician-rated measures of depression severity: can one replace the other? *Depression and Anxiety*, *29*(12), 1043–1049.
- Weiss, A. P., Guidi, J., & Fava, M. (2009). Closing the efficacy-effectiveness gap: translating both the what and the how from randomized controlled trials to clinical practice. *Journal of Clinical Psychiatry*, *70*, 446–449.
- Wood, B. M., Nicholas, M. K., Blyth, F., Asghari, A., & Gibson, S. (2010). The utility of the short version of the depression anxiety stress scales (DASS-21) in elderly patients with persistent pain: does age make a difference? *Pain Medicine*, *11*(12), 1780–1790.
- Young, J. L., Waehler, C. A., Laux, J. M., McDaniel, P. S., & Hilsenroth, M. J. (2003). Four studies extending the utility of the Schwartz outcome scale (SOS-10). *Journal of Personality Assessment*, *80*(2), 130–138.
- Zimmerman, M., & McGlinchey, J. B. (2008). Why don't psychiatrists use scales to measure outcome when treating depressed patients? *Journal of Clinical Psychiatry*, *68*, 1916–1919.