Assessment

Validity of the Generalized Anxiety Disorder-7 Scale in an Acute Psychiatric Sample

Sarah Kertz,* Joe Bigda-Peyton and Throstur Bjorgvinsson
McLean Hospital, Belmont, MA, USA

Generalized anxiety disorder (GAD) is one of the most prevalent psychiatric presentations; however, GAD has the lowest diagnostic reliability of the anxiety disorders and is poorly recognized in clinical practice. A more reliable assessment of GAD could lead to earlier detection and treatment of the disorder, which has an otherwise debilitating course and significant associated impairment. The 7-item GAD Scale (GAD-7) has shown promise as a measure with good clinical utility and strong psychometric properties in primary care and community settings but has yet to be assessed in acute psychiatric populations. This study examined the validity of the GAD-7 in a sample of 232 patients enrolled in a partial hospital programme. Patients completed a diagnostic interview and a battery of self-report measures before and after treatment. Findings suggest that the GAD-7 has good internal consistency and good convergent validity with worry, anxiety, depression and stress, and the measure was sensitive to change over the course of a short intensive cognitive-behavioural therapy partial hospital programme. However, the confirmatory analysis failed to support the hypothesized unidimensional factor structure; and although the GAD-7 demonstrated good sensitivity (.83), specificity was poor (.46) in identifying patients with GAD. Overall, the GAD-7 appears to be a valid measure of generalized anxiety symptoms in this sample, on the basis of good internal consistency, convergent validity and sensitivity to change, but does not perform well as a screener for GAD. Copyright © 2012 John Wiley & Sons, Ltd.

Key Practitioner Message:
• The GAD-7 Scale is an easy-to-score, self-report measure of core generalized anxiety disorder symptoms.
• The GAD-7 Scale has good internal consistency and convergent validity with depression, anxiety, stress and worry, and is sensitive to change.
• The GAD-7 Scale appears to be a good measure of generalized anxiety symptoms in an acute psychiatric sample.
• The GAD-7 Scale does not perform well as a screener for GAD and should not be used to identify cases of GAD in acute psychiatric samples.

Keywords: GAD-7, Generalized Anxiety Disorder, Worry, Psychometric Properties

Generalized anxiety disorder (GAD) is characterized by excessive and uncontrollable worry in addition to three or more associated symptoms including fatigue, difficulty concentrating, sleep trouble, irritability, restlessness and muscle tension (Diagnostic and Statistical Manual of Mental Disorders [DSM], Fourth Edition, Text Revision; American Psychiatric Association, 2004). Lifetime prevalence rates for GAD have been estimated at 5.7% for adults in the USA, with a 12-month prevalence of 3.1% (National Comorbidity Survey-Replication, 2011). GAD presentation is exceedingly common across healthcare settings; in primary care, for instance, prevalence estimates for GAD range from 3.7% to 14.8% (Olfson, 2000; Olfson et al., 1997).

Generalized anxiety disorder is associated with significant impairment and economic burden on the healthcare system. For example, GAD symptom scores have been linked with disability across multiple domains of daily life including self-care, interpersonal functioning and healthcare resource utilization (Ruiz et al., 2011). Furthermore, GAD is associated with an increased healthcare cost of $2138 per individual over the course of a person's lifetime (Marciniak et al., 2005), and estimated health costs for individuals with GAD are 64% higher than those without GAD (Olfson & Gameroff, 2007).

High rates of impairment are likely a function of the debilitating course of GAD. Studies of the trajectory of GAD have yielded inconsistent findings, with results
suggesting that the disorder is either chronic or recurrent (Angst, Gamma, Baldwin, Ajdacic-Gross, & Rossler, 2009; Bruce et al., 2005; Wittchen, 2002). Evidence for the chronicity of the disorder was found in a longitudinal study in which 42% of individuals with GAD (based on DSM-III-R criteria) still had the disorder 12 years later; furthermore, individuals with GAD were symptomatic for 74% of the study’s duration (Bruce et al., 2005). Other work suggests that the disorder is recurrent, with a chronicity rate of 20% (Angst et al., 2009). Regardless of the exact nature of the course, without treatment, the prognosis is poor (Yonkers, Dyck, Warshaw, & Keller, 2000; Yonkers, Warshaw, Massion, & Keller, 1996). Methods for increasing early recognition and appropriate treatment referrals could have wide-ranging benefits.

Unfortunately, identifying GAD appears to be challenging. GAD has the lowest diagnostic reliability of any anxiety disorder (Brown, DiNardo, Lehman, & Campbell, 2001) and is poorly recognized in clinical practice (Beesdo et al., 2009). Estimates suggest that when individuals with GAD seek help for symptoms, only 50–65% are identified as suffering from a psychiatric problem; and of those, only 34% are diagnosed with GAD (Hoyer et al., 1998; Weiller, Bisserbe, Maier, & Recrubier, 1998).

The poor recognition of GAD may be due in part to high rates of comorbidity with other psychiatric disorders and general medical conditions. Comorbidity rates suggest that 90% of individuals from the general population with lifetime GAD have also met criteria for another psychiatric disorder (Carter, Wittchen, Pfister, & Kessler, 2001). Two-thirds of individuals with current GAD also meet criteria for at least one additional psychiatric disorder (Wittchen, Zhao, Kessler, & Eaton, 1994). More specifically, GAD is frequently comorbid with major depressive disorder and other anxiety disorders and has been linked closely with chronic pain conditions, medically unexplained somatic symptoms and sleep disorders (Nutt, Argyropoulos, Hood, & Potokar, 2006).

A practical assessment of GAD could lead to earlier detection and treatment of a disorder with an otherwise debilitating course. Individual distress could be reduced by limiting impairment, disability status and overall healthcare usage; and associated costs of GAD to society could also be lessened. The 7-item GAD Scale (GAD-7; Spitzer, Kroenke, Williams, & Lowe, 2006) has shown promise as a scale with good clinical utility and strong psychometric properties. The measure is a short, self-report questionnaire that is easily scored and was developed specifically to increase recognition of GAD in a primary care patient sample (Spitzer et al., 2006). It has demonstrated good reliability and construct validity, as evidenced by its associations with depression, self-esteem, quality of life, life satisfaction and resilience (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007; Löwe et al., 2008; Spitzer et al., 2006). Importantly, the measure performs well diagnostically and accurately identified 89% of individuals with GAD in a primary care sample (Spitzer et al., 2006).

Although the GAD-7 Scale has promising psychometrics, it has not yet been evaluated in a psychiatric population. The measure’s reliability and validity in psychiatric samples is especially important to understand, as the measure has recently attracted the attention of the National Network of Depression Centers, an organization including 21 leading depression centres and academic medical centres across the nation. The GAD-7 Scale is included in their Common Assessment Package project, an initiative developed to standardize assessment of depression and bipolar diagnosis in both research and clinical settings and is being administered nationwide. The current study evaluates preliminary psychometric properties of the measure, including convergent validity, factor structure, sensitivity, specificity and responsiveness to change, in a naturalistic, partial hospital setting. This setting presents a substantial assessment challenge, as patients are often admitted with complex diagnostic presentations and acute symptom severity.

On the basis of previous work, we hypothesized that the GAD-7 items would demonstrate moderate-to-strong correlations with one another and with the total score and evidence good internal consistency. We predicted that the GAD-7 total score would correlate positively with worry, anxiety and stress, and to a lesser degree with depression, and that patients who meet criteria for GAD would report higher scores than those who do not. We also examined the measure’s factor structure, hypothesizing that all items load onto a unidimensional factor. Sensitivity and specificity were calculated, and it was expected that the measure would show good sensitivity and adequate specificity in identifying patients with GAD. Finally, we predicted that the measure would be sensitive to change over the course of a brief, partial hospitalization.

METHOD

Participants

Participants were 232 patients presenting for treatment at a partial hospital programme in a private psychiatric hospital in New England. About half of patients (47.5%) were referred by their outpatient treatment providers for an increased level of care, whereas the other half (52.5%) were transitioning from an inpatient hospitalization. Specific patient referral information was available for 82% of patients in this sample and indicated that patients referred from inpatient sources were comparable with those from outpatient sources on GAD-7 scores, t(186) = 1.59, p = 0.11, M = 11.38, standard deviation (SD) = 6.57 and M = 12.94, SD = 6.88, respectively. Age of participants ranged from 18 to 68 years, with a mean age of 34.64 years (SD = 13.42), and 60% were women (n = 139). Ethnic composition of the sample was primarily European American (82%; n = 190), followed by Asian (5.6%; n = 13). The majority of the sample...
had never married (59.1%; n = 137). The partial hospital programme serves patients with a wide range of Axis I symptoms. Diagnostic comorbidity in this population is common; and in this sample, 43.1% (n = 100) met criteria for more than one DSM-IV disorder, with an average of 1.56 diagnoses (SD = 0.86). Major depressive disorder is the most common diagnosis (66%; n = 135), followed by GAD (30%; n = 30%), bipolar disorder (19%; n = 40), social anxiety disorder (18%; n = 42), panic disorder (12%; n = 27), obsessive-compulsive disorder (11%; n = 25), mood disorder with psychotic features (11%; n = 25), psychotic disorder (9%; n = 19) and post-traumatic stress disorder (8%; n = 19).

Measures

Miniature International Neuropsychiatric Interview (MINI; Sheehan et al., 1998)

The MINI is a structured diagnostic interview that assesses for Axis I symptoms (e.g., mood, anxiety, alcohol abuse and psychosis). Each MINI diagnostic module consists of a series of screening items followed by questions about specific symptoms. The MINI demonstrates strong reliability and convergent validity with the Structured Clinical Interview for DSM-IV, with inter-rater reliabilities ranging from kappas of 0.89 to 1.0 (Sheehan et al., 1998).

Generalized Anxiety Disorder-7 Scale (Spitzer et al., 2006)

The GAD-7 Scale is a 7-item self-report measure designed to screen for the presence of GAD. Items consist of statements about worry (e.g., ‘Not being able to stop or control worrying’) and general somatic tension (e.g. ‘Trouble relaxing’), which are rated on a 4-point Likert-type scale indicating symptom frequency, ranging from 0 (not at all sure) to 3 (nearly every day). Higher scores indicate higher levels of GAD symptoms. The GAD-7 has demonstrated adequate validity, good clinical utility and generally strong psychometric properties in primary care settings and the general population (Kroenke et al., 2007; Löwe et al., 2008; Spitzer et al., 2006).

Penn State Worry Questionnaire—Abbreviated (PSWQ-A; Hopko et al., 2003)

The PSWQ-A is a single-factor, 8-item measure designed to assess worry severity. Derived from Meyer, Miller, Metzger, and Borkovec’s (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 5-point Likert-type scale ranging from 1 (not at all typical of me) to 5 (very typical of me). Reliability in the present study was very high (α = 0.95).

Schwartz Outcome Scale (SOS; Blais et al., 2008)

The SOS is a well-validated and reliable, single-factor, 10-item measure designed to examine a broad domain of psychological health in a variety of settings (Young, Waehler, Lautz, McDaniel, & Hilsenroth, 2003). Each item assesses for psychological well-being (e.g., ‘My life is according to my expectations’). Participants rate items on a 7-point Likert scale from 0 (never) to 6 (all or nearly all of the time) over the past week. Internal consistency of the SOS was very high in the present study (α = 0.94).

Treatment Setting: Behavioral Health Partil Program (BHPP)

The BHPP is a partial hospitalization program utilizing individual and group cognitive-behavioral therapy to treat a patient population suffering from a variety of Axes I and II personality disorders. Group and individual treatment focuses on the acquisition of cognitive-behavioral therapy skills, with an emphasis on skills related to self-monitoring, cognitive restructuring, behavioral activation and effective communication strategies (for a review of the BHPP treatment approach, see Neuhau, 2006). The average length of stay in the programme is 8.2 (SD = 3.2) days. Patients are assigned a case manager who coordinates all aspects of treatment in conjunction with a programme
psychiatrist who also provides medication management. Treatment includes group and individual psychotherapies.

**Procedure**

Approval for the study was granted by the hospital's Internal Review Board, and all participants gave informed consent. Before receiving any form of treatment at the BHPP, patients completed a demographics survey and a battery of self-report measures. Patients completed the same battery post-treatment. All self-report measures were completed on a computer. Patients were also administered the MINI to assess for current Axis I diagnoses. The MINI was administered by trained psychology interns, doctoral students in psychology and one post-baccalaureate mental health counsellor, who met weekly for supervision with a post-doctoral fellow in clinical psychology.

**Data Analysis**

All GAD-7 data for psychometric analyses were collected pre-treatment, and scores from post-treatment were used only to assess for sensitivity to change. SPSS version 17.0 (IBM Corporation, Armonk, NY) was used for all analyses, with the exception of the factor analyses that used AMOS 17.0, with maximum likelihood estimation.

We examined item characteristics and internal consistency by estimating item means, inter-item correlations and corrected-item total correlations. Internal consistency was estimated using Cronbach's alpha. Construct validity was examined by first assessing convergent validity with correlations between GAD-7 scores and measures of CES-D-10 and DASS—depression, DASS—anxiety, DASS—stress and SOS—overall well-being. Construct validity was also assessed with an examination of GAD-7 means in both the overall sample and a sample of individuals who met criteria for GAD based on the MINI. A factor analysis was conducted to determine if a unidimensional structure fit the data from this sample and tested for invariance across GAD and non-GAD groups. Four criteria were used to examine model fit. First, root mean square error of approximation (RMSEA) was used to assess absolute model fit, with values <.05 considered a 'close fit', values between 0.05 and 0.08 considered 'reasonably close fit' and >.08 an 'unacceptable' model (Steiger, 1990). The Tucker–Lewis Index (TLI), Comparative fit index (CFI) and incremental fit index (IFI) were used to examine relative model fit (compared with the null model), with values >.95 required for a well-fitting model (Bentler, 1990; Hu & Bentler, 1999). The suggested cutoff score of 10 (Spitzer et al., 2006) was used to compare those identified by the GAD-7 as likely having GAD with those who met criteria based on the MINI. Sensitivity, specificity, positive and negative predictive values were also calculated for a range of alternative cutoff scores. Finally, changes in GAD-7 scores over the course of treatment were examined using a repeated measures, mixed analysis of variance design, with GAD diagnosis as the independent factor and time of assessment (before and after treatment) as the between subjects factor.

**RESULTS**

**Sample Characteristics**

The overall mean GAD-7 score in this sample was 10.6 (SD = 5.8), which is in the 97th percentile reported previously in the general population (Löwe et al., 2008) and falls in the moderate anxiety category based on data from primary care samples (Spitzer et al., 2006). Age did not correlate significantly with GAD-7 scores, r = 0.09, p = 0.19. Female participants reported higher scores compared with men, t(225) = -2.75, p = 0.006, M = 13.15 (SD = 6.37) and M = 10.82 (SD = 6.03), respectively.

**Item Characteristics and Internal Consistency**

Means of each item of the GAD-7 and the total score are presented in Table 1. Mean item scores range from 1.0 to 1.9.

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Corrected item-total correlation*</th>
<th>Factor loading</th>
<th>Cronbach's α</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feeling nervous, anxious or on edge</td>
<td>1.9</td>
<td>1.1</td>
<td>0.80</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Not being able to stop or control worrying</td>
<td>1.7</td>
<td>1.1</td>
<td>0.81</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Worrying too much about different things</td>
<td>1.8</td>
<td>1.1</td>
<td>0.79</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Trouble relaxing</td>
<td>1.7</td>
<td>1.1</td>
<td>0.74</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Being so restless that it is hard to sit still</td>
<td>1.0</td>
<td>1.0</td>
<td>0.52</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Becoming easily annoyed or irritable</td>
<td>1.3</td>
<td>1.1</td>
<td>0.53</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Feeling afraid as if something awful might happen</td>
<td>1.2</td>
<td>1.1</td>
<td>0.64</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAD-7 sum score</td>
<td>10.6</td>
<td>5.8</td>
<td>---</td>
<td>---</td>
<td>0.91</td>
</tr>
</tbody>
</table>

GAD-7 = Generalized Anxiety Disorder-7. SD = standard deviation.

*Correlation between the respective item and the total sum score (without the respective item).

Copyright © 2012 John Wiley & Sons, Ltd.  
Item-total correlations ranged from 0.52 to 0.81. Internal consistency was excellent, Cronbach’s α = 0.91 (compared with 0.89, as reported by Löwe et al., 2008). Inter-item correlations ranged from 0.44 to a high of 0.88. Thus, the GAD-7 showed strong internal consistency in this sample.

**Construct Validity**

Convergent validity of the GAD-7 was assessed by examining correlations with CES-D-10, DASS-21, PSWQ-A and SOS. As expected, GAD-7 showed significant positive correlations with PSWQ-A—worry (r = 0.64), DASS—anxiety (r = 0.77) and DASS—stress (r = 0.79) scores. The GAD-7 Scale correlated comparably with two measures of depression, r = 0.64 for DASS—depression and r = 0.74 for CES-D-10—depression. Higher GAD-7 scores were also negatively associated with SOS—overall well-being, r = −0.53. See Table 2 for correlations.

We also compared scores of individuals with and without a diagnosis of GAD (n = 69 and n = 149, respectively, and n = 4 patients did not complete the diagnostic interview). Those who met criteria for GAD reported significantly higher scores, t(169.32) = −4.27, p = 0.001 (Levene’s test was significant, F(221) = 12.16, p = 0.001, so equal variances were not assumed). Individuals with GAD reported a mean score approximately 3 points greater than those without the diagnosis, M = 14.57 (SD = 4.90) and M = 11.17 (SD = 6.60), respectively. Overall, the measure appears to have strong construct validity, as evidenced by good convergent validity and higher scores in patients with a diagnosis of GAD.

A confirmatory factor analysis was conducted to further examine validity. On the basis of previous work (Löwe et al., 2008), the measure was expected to reflect a unidimensional construct. All seven items were set to load on one higher-order GAD factor. Results showed that, overall, the model did not fit the data well, \( \chi^2 = 81.921 \), degrees of freedom (df) = 14, \( p < 0.001 \), CFI = 0.94, TLI = 0.91, IFI = 0.94 and RMSEA = 0.14, with all values outside of the range indicative of a good fit. An examination of the factor loadings suggested that items 5 (‘Being so restless that it is hard to sit still’) and 6 (‘Becoming easily annoyed or irritable’) loaded only moderately onto the latent factor compared with the other items (Table 1).

To determine the areas of misfit in the hypothesized model, we conducted exploratory analyses. Examination of the modification indices suggested that allowing the error terms of items 4 and 5, 5 and 6, and 4 and 6 to covary would improve the model. Associations between these items are consistent with the conceptualization of GAD, as they reflect the associated symptoms criterion of the GAD diagnosis. We respecified the model, allowing the pairs of error variances to covary, and re-estimated the model. The respecified model appeared to fit the data better than the first model, \( \chi^2 = 33.63, \text{df} = 11, \ p < 0.001 \), CFI = 0.98, TLI = 0.96, IFI = 0.98, RMSEA = 0.09, although the RMSEA value suggested an inadequate model. Given that the remaining suggested modifications could not be conceptually justified, the model respecification process was concluded.

We next tested for model invariance across GAD and non-GAD groups. A baseline model with both groups was estimated simultaneously, \( \chi^2 = 58.55, \text{df} = 22, \ CFI = 0.97, \ TLI = 0.94, \ IFI = 0.97, \ RMSEA = 0.09 \), suggesting an overall adequate fit to the data based on CFI and IFI values >0.95. Next, factor loadings, covariances and the variance of the latent GAD factor were constrained to equality, resulting in a \( \chi^2 = 75.6, \text{df} = 32, \ p < 0.001 \). Because the change in chi-square was non-significant, \( \chi^2 \) change = 16.75, df change = 10, \( p = 0.08 \), we conclude that the model is invariant across GAD diagnosis. Overall, a unidimensional factor structure provides a

<table>
<thead>
<tr>
<th>Table 2. Generalized Anxiety Disorder-7 correlations with depression, anxiety, stress, and overall well-being</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>1. GAD-7</td>
</tr>
<tr>
<td>2. CES-D-10—depression</td>
</tr>
<tr>
<td>3. DASS—depression</td>
</tr>
<tr>
<td>4. DASS—anxiety</td>
</tr>
<tr>
<td>5. DASS—stress</td>
</tr>
<tr>
<td>6. PSWQ-A—worry</td>
</tr>
<tr>
<td>7. SOS—overall well-being</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>

All correlations are significant, \( p < 0.001 \). GAD-7 = Generalized Anxiety Disorder-7, CES-D-10 = Center for the Epidemiological Studies of Depression-10, DASS = Depression, Anxiety and Stress Scale-21, PSWQ-A = Penn State Worry Questionnaire—Abbreviated, SOS = Schwartz Outcome Scale, SD = standard deviation.
good fit to the data only when items 4, 5 and 6 are allowed to covary, and the model is comparable across patients with and without a diagnosis of GAD.

**Cutoff Scores, Sensitivity and Specificity**

Sensitivity and specificity values for the GAD-7 Scale were calculated next. Previous work suggests that a cutoff score of 10 maximizes sensitivity and specificity in identifying cases of GAD (Spitzer et al., 2006). In this sample, a cutoff score of 10 resulted in good sensitivity (83%) but poor specificity (46%) and a high false positive rate (0.54). Values were also calculated for a range of other cutoff scores (ranging up to 17), but none adequately balanced sensitivity and specificity. That is, cutoff scores with good sensitivity resulted in poor specificity, whereas those with good specificity resulted in poor sensitivity (see Table 3 for sensitivity, specificity and positive and negative predictive power). A receiver operating characteristic curve analysis estimated the area under the curve at 0.65 (95% confidence interval = 0.59-0.73). See Figure 1.

**Improvements during Treatment**

We also examined changes in GAD-7 scores across the patient's treatment in the programme. A mixed-design, repeated-measures analysis of variance, with GAD diagnosis as the independent factor and time of GAD-7 assessment (admission and discharge) as the between-subjects factor, was conducted. Results showed a significant main effect for time, $F(1, 166) = 84.73, p < 0.001$, partial $\eta^2 = 0.34$, and GAD diagnosis, $F(1, 166) = 8.9, p = 0.003$, partial $\eta^2 = 0.05$, but no interaction between time and GAD diagnosis, $F(1, 166) = 0.29, p = 0.59$. Because the previous t-test indicated that those with GAD report higher GAD-7 scores, the main effect finding here will not be discussed. The effect for time indicated that scores decreased from admission ($M = 12.52$, standard error = 0.54) to discharge ($M = 8.9$, standard error = 0.50), regardless of GAD diagnosis.

**DISCUSSION**

This study examined the validity of the GAD-7 Scale in a sample of patients in an acute partial hospital programme. Patients in this setting frequently have complex diagnostic presentations and comorbidity is the norm, rather than the exception. Although the GAD-7 Scale has demonstrated strong psychometric properties in primary care (Spitzer et al., 2006) and the general population (Löwe et al., 2008), it is yet to be examined in a psychiatric population.

In this sample, GAD-7 scores were uniformly high across the sample. The average GAD-7 score fell in the 97th percentile, compared with individuals from the general population (Löwe et al., 2008). Those who met diagnostic criteria for GAD reported higher scores than those who do not.

**Table 3. Sensitivity, specificity and false positive rates for possible Generalized Anxiety Disorder-7 cutoff scores**

<table>
<thead>
<tr>
<th>GAD-7 cutoff score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False positive rate</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.83</td>
<td>0.46</td>
<td>0.54</td>
<td>0.40</td>
<td>0.86</td>
</tr>
<tr>
<td>11</td>
<td>0.77</td>
<td>0.50</td>
<td>0.50</td>
<td>0.40</td>
<td>0.83</td>
</tr>
<tr>
<td>12</td>
<td>0.74</td>
<td>0.55</td>
<td>0.45</td>
<td>0.41</td>
<td>0.83</td>
</tr>
<tr>
<td>13</td>
<td>0.64</td>
<td>0.57</td>
<td>0.43</td>
<td>0.39</td>
<td>0.78</td>
</tr>
<tr>
<td>14</td>
<td>0.59</td>
<td>0.62</td>
<td>0.38</td>
<td>0.41</td>
<td>0.78</td>
</tr>
<tr>
<td>15</td>
<td>0.54</td>
<td>0.69</td>
<td>0.31</td>
<td>0.43</td>
<td>0.77</td>
</tr>
<tr>
<td>16</td>
<td>0.48</td>
<td>0.73</td>
<td>0.27</td>
<td>0.43</td>
<td>0.76</td>
</tr>
<tr>
<td>17</td>
<td>0.41</td>
<td>0.76</td>
<td>0.24</td>
<td>0.42</td>
<td>0.75</td>
</tr>
</tbody>
</table>

PPV = positive predictive value; NPV = negative predictive value.
not, and the GAD group mean is nearly identical to that found in previous work (Spitzer et al., 2006); however, patients without GAD also reported high scores. The average score in patients who do not meet criteria for GAD falls in the 97th percentile on the basis of norms from the general population (Löwe et al., 2008). Further, both group scores fell above the suggested cutoff score for identifying individuals with GAD. Given that all patients had been admitted to the programme at the time of assessment, it seems unlikely that non-GAD patients are over-reporting symptoms in order to access services.

Our hypothesis that the GAD-7 Scale would have good internal consistency and item-total correlations was supported. The GAD-7 Scale displayed excellent internal consistency in this sample. Cronbach’s alpha was excellent, and inter-item correlations ranged from moderate to high. Item-total correlations also ranged from moderate to high, with the ‘restless’ and ‘irritable’ items correlating the lowest of the seven items. Similarly, these items showed the weakest item-total correlations in a study of the general population, (Löwe et al., 2008).

The hypothesis that the measure would show good convergent validity was largely supported. The GAD-7 correlated highly with DASS—anxiety and DASS—stress scores, comparable with previous work (Löwe et al., 2008; Spitzer et al., 2006). The measures also showed a moderate negative association with SOS—overall well-being scores, in line with our prediction. Of note, the GAD-7 also correlated strongly with both DASS—depression and CES-D-10—depression scores and only moderately with PSWQ-A—worry scores. GAD and depression are often comorbid (Hettema, 2008), and other work suggests that worry levels are comparable in individuals with GAD and depression (Starcevic, 1995). Therefore, the correlation of the GAD-7 with depression measures is not surprising. Future work might strive to better understand if the GAD-7 is a measure of GAD symptoms specifically or negative affect more generally.

A confirmatory factor analysis, specified so that each of the seven items loaded onto one latent GAD factor, failed to support our hypothesis and the model detailed elsewhere (Löwe et al., 2008). Follow-up exploratory analyses suggested that several of the items, specifically items 4, 5 and 6, shared unique variance beyond that explained by the GAD factor. An examination of the items suggests that perhaps these items also reflect a somatic tension/autonomic arousal factor. A recent review has suggested that it may be beneficial to examine the subtypes of GAD (Portman, Starcevic, & Beck, 2011), on the basis of findings that individuals who would otherwise meet diagnostic criteria for GAD (based on endorsement of the associated symptoms criterion) fail to receive the diagnosis because they do not meet the excessive worry criterion (Roth et al., 2008). Portman et al. (2011) hypothesize that there may be subtypes of GAD, including an excessive worry type, a somatic tension/autonomic arousal subtype and a combined subtype. To date, no work has tested this hypothesis empirically, although our findings here suggest that this may be valuable to explore in future work.

The hypothesis that the GAD-7 Scale would show good sensitivity was supported, whereas the prediction of good specificity was not. High scores on the GAD-7 across participants, as discussed previously, impacted the measure’s utility to discriminate those who met criteria for GAD from those who did not. Previous research suggested that a cutoff score of 10 resulted in good sensitivity and specificity in a primary care sample, with sensitivity of 89% and specificity of 82% (Spitzer et al., 2006). Contrary to these findings, however, a cutoff score of 10 in this sample resulted in good sensitivity (0.83) but poor specificity (0.46). Further, the cutoff score resulted in an unacceptable rate of false positives.

We examined the potential utility of a range of other cutoff scores, up to 17, to determine if a higher score might perform better in our more symptomatic sample; however, the alternative cutoff scores were similarly limited and unable to balance sensitivity and specificity so that both were acceptable.

Findings from this study indicate that the GAD-7 Scale is sensitive to changes in symptoms over time, in line with the hypothesis. There was no interaction between time of assessment and GAD group, which suggests that all patients saw symptom improvement over time, regardless of GAD diagnosis. Given the chronic nature of GAD (Bruce et al., 2005), the ability of the GAD-7 Scale to detect symptom improvement over a short period is a significant strength. The main effect for time indicates that patient scores decreased approximately 3.5 points, from a mean score of 12.5 to 9.9. This change suggests that, on average, patients are moving across the previously identified clinical cutoff point of the GAD-7, from a group identified as likely having GAD to a non-GAD group; however, this particular finding should be interpreted with caution, given the poor specificity of the cutoff score in this sample.

This study has several strengths. It is the first to examine the psychometric properties of the GAD-7 in a psychiatric sample and the diagnostically heterogeneous nature of the sample has strong external validity, suggesting that findings are likely to be highly generalizable. The inclusion of a range of symptom measures, including depression, anxiety, worry, stress and overall well-being scores, also allowed for a thorough assessment of convergent validity. Despite these strengths, however, the study is not without limitations. First, the acute nature of the symptoms in the current sample could have influenced the results, which may differ in outpatient clinics. Future work would benefit from examining the psychometric properties of the GAD-7 in a traditional outpatient setting. In addition, the sample was relatively homogenous in terms of ethnic and racial demographics, and these findings would need to be replicated in more diverse samples.
Overall, the GAD-7 Scale is a short, easily administered and scored self-report questionnaire with strong construct validity in this acutely symptomatic psychiatric sample, as evidenced by good internal consistency, good convergent validity and higher scores in patients with GAD compared with those without GAD. It should be noted that the GAD-7 Scale also correlated highly with DASS and CES-D-10—depression scores and moderately with FSWQ-A—worry scores. The measure was responsive to symptom improvement over the course of treatment in a partial hospitalization programme and may be a valuable indicator of fluctuations in anxiety over time. However, the initial hypothesized factor structure failed to provide a good fit to the data; and the GAD and non-GAD groups reported scores higher than the suggested cutoff score, which resulted in good sensitivity but poor specificity. In summary, the GAD-7 is an internally consistent and valid measure of general anxiety and worry in a sample of acutely symptomatic patients, but the measure is unlikely to be useful as a screening tool for GAD in this sample.

REFERENCES