



Reliability and validity of the dimensional features of generalized anxiety disorder[☆]



Lauren A. Rutter^{*}, Timothy A. Brown

Center for Anxiety and Related Disorders, Boston University, United States

ARTICLE INFO

Article history:

Received 27 March 2014
Received in revised form 23 July 2014
Accepted 20 October 2014
Available online 30 October 2014

Keywords:

GAD
Reliability
Validity
Dimensional classification
Diagnostic criteria

ABSTRACT

The reliability and validity of the dimensional features of generalized anxiety disorder (GAD) were examined in a diverse sample of 508 outpatients with anxiety and mood disorders who underwent two independent administrations of the Anxiety Disorders Interview Schedule for *DSM-IV*: Lifetime version (ADIS-IV-L; Di Nardo, Brown, & Barlow, 1994). Inter-rater reliability was higher in the full sample than in patients with current GAD. Additionally, the presence of a mood disorder weakened inter-rater reliability. We also explored the unique contribution of excessiveness and uncontrollability of worry to various clinical outcomes and found that excessiveness predicted anxiety, depression, and stress self-report measures, and uncontrollability predicted clinical severity and number of diagnoses. Findings are discussed with regard to their implications for the classification of GAD (e.g., utility of dimension-based assessment to improve the classification of psychological disorders).

© 2014 Elsevier Ltd. All rights reserved.

Classification of anxiety and mood disorders has been an inexact science, reflected by the modest reliability of many diagnostic categories (e.g., Brown, Di Nardo, Lehman, & Campbell, 2001). Generalized anxiety disorder (GAD) is a diagnostic category that has undergone substantial revisions (Brown, Barlow, & Liebowitz, 1994). GAD is classified in the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric Association, 2013) as chronic (lasting at least six months), excessive anxiety and worry about a number of events or activities that is difficult to control, and is associated with at least three of six symptoms of tension/negative affect with some present more days than not for at least six months. For GAD to be assigned, the worries and associated symptoms must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. Additionally, a diagnosis of GAD requires that the anxiety and worry do not occur exclusively during the course of a mood or psychotic disorder.

When GAD first appeared in *DSM-III* (American Psychiatric Association, 1980), it was a residual category, diagnosed only if a patient did not meet criteria for any other anxiety or mood

disorder. This definition was associated with low inter-rater reliability ($\kappa = .47$; Di Nardo, O'Brien, Barlow, Waddell, & Blanchard, 1983). The reformulation of GAD in *DSM-III-R* (American Psychiatric Association, 1987) failed to considerably improve the reliability of the disorder, as was shown by large-scale studies entailing administration of two independent structured interviews (κ s for current GAD were .27 in Mannuzza, Fyer, Martin, & Gallops, 1989; .53 in Di Nardo, Moras, Barlow, & Rapee, 1993; and .56 in Williams et al., 1992). Evidence of low reliability and high comorbidity of GAD with other disorders (comorbidity rates exceeding 80%; see Brown & Barlow, 1992) led researchers to question whether there was sufficient discriminant validity to retain GAD as a diagnostic category in *DSM-IV* (Brown et al., 1994). The diagnostic criteria were revised substantially in *DSM-IV* (American Psychiatric Association, 1994) in an effort to define the boundary of GAD with other anxiety disorders, mood disorders, adjustment disorders, and nonpathological worry. Revisions to *DSM-IV* included the requirement that worry must be perceived as uncontrollable (based on evidence that uncontrollability of worry distinguishes GAD worry from normal worry; Abel & Borkovec, 1995; Borkovec & Roemer, 1994).

Another substantial change in *DSM-IV* was the reduction of associated symptoms from 18 to 6. Symptoms of autonomic arousal (e.g., accelerated heart rate, shortness of breath) were eliminated, while symptoms of tension and negative affect (e.g., muscle tension, irritability) were retained. Although this change was partly data-driven (e.g., Brown, Marten, & Barlow, 1995), researchers were concerned that this revision would further obfuscate the boundary between GAD and mood disorders (Clark & Watson, 1991). Nonetheless, the revisions to *DSM-IV* GAD diagnostic criteria were

[☆] This study was supported by Grant MH039096 from the National Institute of Mental Health to the second author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

^{*} Corresponding author at: Center for Anxiety and Related Disorders, Department of Psychology, Boston University, 648 Beacon Street, 6th Floor, Boston, MA 02215-2013, United States. Tel.: +1 617 353 9610; fax: +1 617 353 9609.

E-mail address: lrutter@bu.edu (L.A. Rutter).

associated with increased diagnostic reliability ($\kappa = .67$ in Brown et al., 2001), compared to *DSM-III-R* ($\kappa = .53$ in Di Nardo et al., 1993).

In addition to examining the diagnostic reliability of the various *DSM-IV* anxiety and mood disorders, Brown et al. (2001) evaluated the factors most commonly involved in diagnostic disagreements; e.g., difference in patient report, threshold disagreements (e.g., difficulties applying cutoffs for presence or absence of a disorder based on sufficient distress or impairment), change in clinical status, interviewer error, and diagnosis subsumed under comorbid condition. With regard to the diagnosis of GAD, difference in patient report was the most common source of disagreement (55%). A reliable diagnosis of GAD calls for consistent self-report of many subjective features, their onset, and their duration in relation to other conditions (e.g., mood disorders). Inconsistency in patient reports could be indicative of vagueness of these diagnostic features and patients' difficulty distinguishing them from other disorders (Brown et al., 2001). This study also found that GAD diagnostic disagreements involved mood disorders in 47% of cases, which is consistent with prior evidence that boundary issues with mood disorders pose a larger problem for GAD than do other anxiety disorders (e.g., Brown, Chorpita, & Barlow, 1998).

In addition to work on reliability at the diagnostic level, researchers have begun to explore the reliability of the dimensional features of GAD. Gordon and Heimberg (2011) examined the reliability of GAD features in a sample of 129 patients with a principal diagnosis of GAD. GAD features were assessed using the Anxiety Disorders Interview Schedule for *DSM-IV*: Lifetime version (ADIS-IV-L; Di Nardo, Brown, & Barlow, 1994). As estimated by intraclass correlations (ICCs), the dimensions of excessiveness of worry, uncontrollability of worry, and interference due to worry were found to have fair to good inter-rater agreement (i.e., ICCs = .60, .59, and .62, respectively). Agreement was poor for distress due to worry (ICC = .30). Agreement for the associated symptoms ranged from poor to good (range of ICCs = .22–.65), and varied by symptom (i.e., good for fatigue, fair for irritability, muscle tension, sleep disturbance, and concentration difficulties, and poor for restlessness) (Gordon & Heimberg, 2011).

In addition, Gordon and Heimberg (2011) investigated the convergent and discriminant validity of dimensional features of GAD. Evidence for convergent validity of the clinical features of GAD with a questionnaire measure of pathological worry was somewhat modest. Specifically, although all correlations were statistically significant ($ps < .05$), the ADIS-IV-L rated dimensions of GAD were weakly correlated with the Penn State Worry Questionnaire (PSWQ); i.e., excessiveness ($r = .27$), uncontrollability ($r = .26$), distress due to worry ($r = .32$) and the clinical severity of GAD ($r = .20$). Moreover, some features of GAD evidenced poor discriminant validity with social anxiety (e.g., correlation of excessiveness of worry with Social Anxiety Interaction Scale = .46, $p < .001$). It is possible that reliability and concurrent validity estimates were attenuated by range restriction (i.e., less variability in measures of GAD and worry) because the sample was limited to outpatients with a principal diagnosis of GAD. In addition, Gordon and Heimberg (2011) did not assess disorders other than GAD and social phobia (i.e., the second interviewer did not administer the entire ADIS-IV-L), which may also have adversely impacted the reliability estimates. Administering solely the GAD and social phobia modules of the ADIS-IV-L may have resulted in symptoms being incorrectly attributed to GAD and social phobia, when the symptoms were actually due to another disorder that was not fully assessed, such as a mood disorder.

Another issue bearing on the validity of the dimensional features of GAD pertains to the distinctiveness of the diagnostic criteria of excessiveness and uncontrollability of worry. Some researchers (e.g., Andrews et al., 2010) have suggested that the excessiveness and uncontrollability criteria are redundant due to their

conceptual similarity and strong association (e.g., $r = .91$; Brown et al., 2001). Moreover, a study has shown that only about 4% of respondents who met criteria for GAD reported that their worries were excessive, but still controllable (Beesdo et al., 2011). Even though excessiveness and uncontrollability were retained as key, separate features for GAD in *DSM-5* (i.e., the criteria for GAD did not change between *DSM-IV* and *DSM-5*), the question remains about their discriminant validity. To examine the incremental validity of the uncontrollability criterion, Hallion and Ruscio (2013) recently conducted a study of 126 adults with GAD drawn from a community sample. Although the two features were intercorrelated highly ($r = .83$), it was found that uncontrollability incrementally added to the prediction of GAD and more general measures pertaining to clinical severity, comorbidity, and treatment-seeking after holding excessiveness constant. In most analyses, excessiveness was no longer a significant predictor after uncontrollability entered the model (Hallion & Ruscio, 2013).

The present study adds to previous work examining reliability and validity of GAD features (Gordon & Heimberg, 2011; Hallion & Ruscio, 2013), through the use of (a) a larger sample of outpatients who underwent two, independent administrations of the ADIS-IV-L, (b) double administrations of the complete ADIS-IV-L (not GAD and social phobia modules only as in Gordon & Heimberg, 2011), (c) second interviewers who were not cognizant of the first interviewer's ADIS-IV-L diagnoses, and (d) a more diverse clinical sample (i.e., the sample was not limited to GAD cases only). Because range restriction typically attenuates correlational estimates (e.g., reliability), we expected to find greater inter-rater reliability of GAD features in the full sample compared to GAD cases only. Furthermore, we evaluated four potentially salient moderators of inter-rater reliability: (1) days separating ADIS-IV-L assessments, (2) presence of a mood disorder, (3) number of diagnoses assigned, and (4) severity of worry (as indicated by the PSWQ scores). We predicted inter-rater reliability to decrease as the number of diagnoses increases, and we expected reliability to decrease as number of days between assessments increases. We predicted that the presence of a mood disorder would weaken reliability (due to well-known boundary issues between GAD and mood disorders described above). We also predicted that severity would impact reliability, with more severe GAD leading to higher reliability. Finally, we explored the question of whether both excessiveness and uncontrollability uniquely contribute to the prediction of overall GAD severity, number of diagnoses, negative affect, and depressive symptoms.

1. Method

1.1. Participants

Participants were 508 patients who presented for assessment and treatment at the Center for Anxiety and Related Disorders (CARD) at Boston University between March 1997 and August 2012. Individuals underwent a semi-structured interview and completed a series of self-report questionnaires at the time of their initial assessment. The sample was randomly selected from a larger sample to receive two independent administrations of the ADIS-IV-L. Women constituted a larger portion of the sample (59%). The average age was 31.82 ($SD = 10.32$, range 18–66). The sample was predominantly Caucasian (89%; African-American = 3%; Asian = 4%; Other = 4%). Participants were required to be age 18 or older and to have a presenting complaint involving anxiety or mood symptoms. Participants were excluded from the study if any of the following were present: (1) current delusions or hallucinations, (2) current suicidal or homicidal risk meriting intervention, or (3) two or more hospitalizations in the last 5 years for psychotic symptoms. Patients

were also required to meet stabilization/wash-out criteria for psychotropic medications and psychotherapy for the time periods preceding and overlapping their diagnostic assessment (see Brown et al., 2001, for details). The study procedures were approved by the Institutional Review Board of Boston University.

Diagnoses were established using the ADIS-IV-L, a semi-structured interview designed to ascertain reliable diagnosis of the DSM-IV anxiety, mood, somatoform, and substance use disorders, and to screen for the presence of other conditions (e.g., psychotic disorders). The ADIS-IV-L has been shown to have good or excellent diagnostic reliability for most anxiety disorders (see Brown et al., 2001). The second interview was completed before treatment was initiated, with an average of 9.49 days after the first interview ($SD=8.47$, range 0–66). After both interviews were complete and interviewers had independently scored them, cases were presented at weekly staffing meetings. This entailed a presentation of interviewers' diagnoses, discussion of factors contributing to diagnostic disagreements, and establishment of consensus diagnoses.

For each diagnosis, interviewers assigned a 0–8 dimensional clinical severity rating (CSR) to indicate the degree of distress and impairment associated with the disorder (0 = “none” to 8 = “very severely disturbing/disabling”). For current and lifetime disorders that meet or surpass the threshold for a formal DSM-IV diagnosis, CSRs of 4 (definitely disturbing/disabling) or higher are assigned (“clinical” diagnoses). For patients with two or more current clinical diagnoses, the “principal” diagnosis is the one receiving the highest CSR, and “additional” diagnoses refer to current clinical diagnoses that are not principal. CSRs of 3 and below (subclinical diagnoses) were assigned when the key features of a current or lifetime disorder existed but were not judged to be extensive or severe enough to merit a formal DSM-IV diagnosis.

The rates of current clinical disorders (either principal or additional) that frequently occurred in the sample were as follows: panic disorder with or without agoraphobia (PD/A) (28.1%), social phobia (SOC) (41.9%), GAD (24.4%, with 9.5% having GAD as their principal diagnosis), obsessive-compulsive disorder (OCD) (15.4%), posttraumatic stress disorder (PTSD) (3.5%), specific phobia (SPEC) (18.9%), anxiety disorder not otherwise specified (9.3%), major depressive disorder (MDD) (28.0%), dysthymia (DYS) (8.1%), and depressive disorder not otherwise specified (4.1%).

1.2. Measures

1.2.1. Anxiety Disorders Interview Schedule for DSM-IV: Lifetime version (ADIS-IV-L; Di Nardo et al., 1994)

The ADIS-IV-L provides an assessment of lifetime disorders and a diagnostic timeline that allows for accurate determination of onset, remission, and temporal sequence of current and lifetime disorders. The ADIS-IV-L dimensionally assesses key and associated features of anxiety and mood disorders including SOC, GAD, OCD, SPEC, MDD, and DYS; these features are dimensionally rated regardless of whether a formal DSM-IV diagnosis is under consideration. Interviewers assigned dimensional ratings (0–8 scales) for ratings of excessiveness of worry and uncontrollability of worry in eight worry domains, and the six symptoms comprising its associated symptoms criterion. Interviewers followed appropriate DSM-IV duration criteria while making dimensional ratings (i.e., ratings reflected a composite of severity, frequency, or duration in respect to the DSM-IV criterion, if specified).

Ratings within the GAD section of the ADIS-IV-L were used to compute dimensional composite scores for GAD features. For excessiveness and uncontrollability of worry, the 0–8 dimensional ratings for excessiveness (0 = “no worry/no tension” to 8 = “constantly worried/extreme tension”) and uncontrollability (0 = “never/no difficulty” to 8 = “constantly/extreme difficulty”) for

each of the eight worry domains (i.e., minor matters, work/school, family, finances, social/interpersonal, health (self), health (significant others), and community/world affairs) were summed to create a composite score for each. A composite score for associated symptoms was created by summing the 0–8 dimensional ratings (0 = “none” to 8 = “very severe”) for the following: restlessness, fatigue, difficulty concentrating, irritability, muscle tension, and sleep difficulties.

After completing the ADIS-IV-L, interviewers assigned 0–8 ratings (0 = “none” to 8 = “very severely disturbing/disabling”) for the each of the DSM-IV criteria of GAD. From these ratings, a composite score was created by summing the following: (a) excessive anxiety and worry, (b) uncontrollability of worry, (c) frequency and intensity of the six associated symptoms, and (d) interference and distress due to worry and associated symptoms.

The ADIS-IV-L interviews were administered by clinical psychology doctoral students and doctoral-level clinical psychologists. Before participating in the study, interviewers underwent extensive training to meet strict certification criteria in administration of the ADIS-IV-L. Training began with trainees reading the ADIS-IV-L manual, observing interviews, and then conducting collaborative interviews. While observing and collaborating on interviews, the trainees made ratings and diagnoses to compare with the trainer. In the collaborative phase, trainees assumed primary responsibility for the ADIS-IV-L administration, with senior interviewers interjecting as needed. During the certification phase of training, trainees administered a minimum of three ADIS-IV-L interviews under observation of a senior interviewer. After completing the interview, the trainee and trainer independently assigned current and lifetime diagnoses. The matching criteria were as follows: within three of five consecutive interviews, the trainee's clinical diagnoses had to match the trainer's clinical diagnoses within one CSR, and the trainee could not commit any ADIS-IV-L administration errors based on a checklist of nine items (cf. Brown et al., 2001).

1.2.2. Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990)

The Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) consists of 16 items rated on a 5-point Likert-type scale (1 = “not at all typical of me” to 5 = “very typical of me”) and assesses the extent to which worry is pervasive, excessive, and difficult to control (e.g., “I am always worrying about something”). The PSWQ has been shown to have excellent internal consistency ($\alpha = .91$; Meyer et al., 1990) and good convergent and discriminant validity for GAD compared to other anxiety disorders and community controls (Brown, Antony, & Barlow, 1992).

1.2.3. Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995)

The Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995) consists of 21 items rated on a 4-point Likert-type scale (0 = “did not apply to me at all” to 3 = “applied to me very much, or most of the time”) and assesses levels of depression, general anxiety, and general tension/negative affect symptoms. The factor structure of the DASS has been substantiated by both exploratory and confirmatory factor analyses (e.g., Brown, Chorpita, Korotitsch, & Barlow, 1997; Lovibond & Lovibond, 1995).

1.3. Data analysis

SPSS 20.0 was used to conduct the analyses. Inter-rater reliability was calculated using intraclass correlation coefficients (ICCs). Inferential tests of the differential magnitude of ICCs were used to test the hypothesized moderators of inter-rater reliability. These tests were computed using the procedures outlined in McGraw and

Table 1
Intraclass correlation coefficients for GAD features in the full sample ($N=508$) and GAD only cases ($n=124$).

Measure	ICC	
	Full sample	GAD only
Excessiveness of worry	.72	.43
Uncontrollability of worry	.77	.54
Associated symptoms	.79	.56
DSM-IV GAD criteria ratings	.81	.60

Note. ICC, intraclass correlation coefficient; GAD, generalized anxiety disorder; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th edition.

Wong (1996, see Table 8, equation for Case 1, one-way random). Multiple regressions were conducted to examine the unique contribution of excessiveness and uncontrollability of worry in predicting selected GAD-relevant clinical outcomes.

2. Results

2.1. Inter-rater reliability

ICCs were interpreted in a manner similar to kappa coefficients, based on guidelines used in previous studies using kappa coefficients to assess reliability of anxiety and mood disorders (e.g., Brown et al., 2001; Gordon & Heimberg, 2011). According to these reliability standards, excellent agreement is indicated by $ICC \geq .75$, good agreement is indicated by $.60 \leq ICC \leq .74$, fair agreement is indicated by $.40 \leq ICC \leq .59$, and poor agreement is indicated by $ICC < .40$.

In Table 1, we present reliability findings for GAD features. Consistent with prediction, reliability of GAD features was higher in the full sample ($N=508$) than in GAD cases only ($n=124$). For example, the ICC for excessiveness of worry was .72 for the full sample, and .43 for GAD cases only. Using the aforementioned standards, we found that uncontrollability of worry, associated symptoms, and DSM-IV GAD criteria ratings showed excellent agreement for the full sample (range of ICCs = .77–.81). Excessiveness of worry evidenced good agreement for the full sample ($ICC = .72$). For GAD only cases, agreement was fair for excessiveness, uncontrollability, and associated symptoms (range of ICCs = .43–.56), and agreement was good for the DSM-IV GAD criteria ratings ($ICC = .60$).

2.2. Moderators of inter-rater reliability

To evaluate the impact of (1) a comorbid mood disorder, (2) days between assessments, (3) GAD severity, and (4) number of diagnoses on reliability, a test of the differential magnitude of ICCs was used with the DSM-IV GAD dimensional score as the outcome variable (cf. McGraw & Wong, 1996). These analyses utilized the full sample and required dichotomous moderator variables. For presence of a comorbid mood disorder, reliability was significantly better for non-depressed cases ($ICC = .78$) than cases with comorbid depression ($ICC = .69$), $F = 1.50$, $p < .001$. For days separating assessments, reliability was slightly better when days between assessments was less than a week ($ICC = .83$) than when assessments were a week or more apart ($ICC = .79$), but the differential magnitude of these ICCs only approached significance, $F = 1.25$, $p = .06$. However, when the cutoff was extended to two weeks, the differential magnitude of ICCs was significant, $F = 1.59$, $p < .001$; i.e., reliability was significantly better when the days separating assessments was two weeks or less ($ICC = .82$) compared to more than two weeks ($ICC = .73$). Higher and lower worry severity was indexed by the sample median of PSWQ scores ($Mdn = 64$). The ICCs were quite similar in magnitude and did not significantly differ as a function of worry severity;

ICCs = .73 and .70 for lower and higher worry severity, respectively, $F = 1.12$, $p = .20$. In a direction counter to prediction, reliability was higher if the number of current diagnoses assigned was two or more ($ICC = .76$) than if number of diagnoses was one ($ICC = .70$), and this difference was statistically significant, $F = 1.29$, $p = .02$.

2.3. Unique contribution of excessiveness and uncontrollability of worry to clinical outcomes

We conducted multiple regression analyses to examine the unique contribution of excessiveness and uncontrollability to the prediction of (1) GAD CSR, (2) PSWQ, (3) number of diagnoses other than GAD, and (4) depression and negative affect/stress symptoms. The correlation between excessiveness and uncontrollability was $.90$, $p < .001$. Regression results are presented in Table 2. For the GAD CSR, uncontrollability contributed to prediction at the $p < .001$ level, but excessiveness did not. PSWQ scores were predicted by both excessiveness and uncontrollability of worry. Number of diagnoses other than GAD was predicted by uncontrollability of worry, but not excessiveness. Stress (negative affect) and depression symptoms, as measured by the DASS, were predicted solely by excessiveness of worry, at the $p < .01$ and $p < .05$ levels, respectively. The effect sizes (f^2) for the full regression models and individual predictors are also presented in Table 2. Although the effect sizes for the regression models were medium to large per Cohen's (1988) standards (range of f^2 s = .12–.52), the effect sizes for the individual predictors that attained statistical significance were uniformly small (range of f^2 s = .01–.04). The differential magnitude of these effect sizes (i.e., full model vs. individual predictors) indicates that the majority of the predictive utility of excessiveness and uncontrollability stems from the variability shared by these features.

3. Discussion

The current study examined the reliability and validity of DSM-IV GAD features in a sample of outpatients diagnosed with a range of anxiety and mood disorders. We explored inter-rater agreement, moderators of reliability, and the unique contribution of excessiveness and uncontrollability in predicting variables such as GAD severity, negative affect, depression symptoms, and number of diagnoses.

Our research adds to Gordon and Heimberg's (2011) work by utilizing a larger clinical sample and a more thorough assessment (full ADIS-IV-L), with interviewers blind to each other's diagnoses and dimensional ratings. As noted earlier, Gordon and Heimberg's (2011) estimates may have been attenuated by range restriction through using a GAD-only sample. In accord with this interpretation, the current results show considerably higher reliability in the full sample than in GAD-only cases. For example, whereas Gordon and Heimberg found the ICC for excessiveness of worry to be .60, we found it to be .43 in the GAD only sample, and .72 in the full sample. Gordon and Heimberg's ICC for uncontrollability of worry was .59, and we obtained ICCs of .54 in the GAD only sample and .77 in the full sample.

In addition, the range of days between assessments was considerably shorter in our study than in Gordon and Heimberg (2011) where the time between assessments in that study spanned from 3 to 260 days, with four patients having 200 or more days between assessments. Our findings suggested that the length of time between diagnostic interviews does have a significant impact on inter-rater reliability; i.e., inter-rater agreement decreases as the number of days between assessments increase. Whereas the differential magnitude of the ICCs approached significance ($p = .06$) when a one-week cut-off was used, inter-rater reliability was

Table 2

Regression models of excessiveness and uncontrollability of worry predicting clinical severity, pervasive worry, number of diagnoses, stress/negative affect, and depression.

Variable	β	<i>B</i>	SE _B	<i>t</i>	<i>F</i>
GAD clinical severity rating (model $R^2 = .23$; $f^2 = .30$)					
Excessiveness of worry ($f^2 = .005$)	.14	.03	.02	1.51	75.70***
Uncontrollability of worry ($f^2 = .03$)	.36	.08	.02	3.96***	
Number of diagnoses other than GAD (model $R^2 = .14$; $f^2 = .16$)					
Excessiveness of worry ($f^2 = .004$)	.13	.01	.01	1.34	39.65***
Uncontrollability of worry ($f^2 = .01$)	.25	.03	.10	2.60**	
PSWQ (model $R^2 = .34$; $f^2 = .52$)					
Excessiveness of worry ($f^2 = .02$)	.36	.48	.11	4.27***	123.40***
Uncontrollability of worry ($f^2 = .04$)	.24	.31	.11	2.78**	
DASS Stress Scale (model $R^2 = .15$; $f^2 = .18$)					
Excessiveness of worry ($f^2 = .02$)	.27	.25	.09	2.88**	42.97***
Uncontrollability of worry ($f^2 = .003$)	.12	.11	.09	1.28	
DASS Depression Scale (model $R^2 = .11$; $f^2 = .12$)					
Excessiveness of worry ($f^2 = .01$)	.21	.19	.09	2.17*	29.19***
Uncontrollability of worry ($f^2 = .003$)	.12	.11	.09	1.26	

Note. GAD, generalized anxiety disorder; PSWQ, Penn State Worry Questionnaire; DASS, Depression Anxiety Stress Scales.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

significantly better ($p < .001$) when the double interviews were conducted within a two-week timeframe. As noted earlier, the most common source of diagnostic disagreements for GAD is difference in patients' reports between interviews (Brown et al., 2001). It is likely that inconsistencies in patient reporting are exacerbated by longer separations between interviews. Another potential source of unreliability that is directly related to the length between interviews is change in clinical status (e.g., a shift from meeting DSM-IV diagnostic threshold to partial remission between interviews). However, given the chronic nature of GAD, clinical status change may be a less common source of unreliability than for conditions characterized by a fluctuating course (e.g., MDD, PD/A). Although longer time spans between interviews are likely to have a deleterious effect on the inter-rater reliability of all disorders, this impact may be most pronounced for episodic conditions.

Our findings add to the extant literature through our inclusion of comorbid mood disorders and the examination other important and previously understudied moderators such as days separating assessments, GAD severity, and number of diagnoses. The finding that the presence of comorbid mood disorders negatively impacted the inter-rater reliability of GAD had previously been observed descriptively at the categorical level (Brown et al., 2001), but to our knowledge, has not been examined statistically using GAD dimensional features. The fact that mood disorder comorbidity had a deleterious effect on GAD reliability may be due to a common source of unreliability at the diagnostic level: disagreements between interviewers about whether the patient's symptoms are reflective of GAD versus a mood disorder.

We conducted multiple regressions to examine the incremental validity of the excessiveness and uncontrollability criteria. In accord with previous studies (e.g., Brown et al., 2001; Hallion & Ruscio, 2013), the interview ratings of excessiveness and uncontrollability of worry were highly correlated ($r = .90$), indicative of questionable discriminant validity. Nevertheless, our results indicated that these two features of worry each significantly predicted outcome variables, but in different ways. A pattern of differential relationships arose from the regression models: when uncontrollability was a significant predictor, it most often uniquely predicted a clinical rating outcome (CSR and number of diagnoses), while excessiveness uniquely predicted questionnaires only (PSWQ and DASS scores). With regard to the effect on the CSR outcome, it may

be that clinicians emphasize uncontrollability over excessiveness in rating the severity of GAD. Both excessiveness and uncontrollability contributed to the prediction of the PSWQ (a questionnaire measure of trait worry), but the effect of excessiveness was stronger. This may be because more items on the PSWQ seem to measure excessiveness of worry (e.g., 15: "I worry all the time") than uncontrollability of worry (e.g., 14: "Once I start worrying, I can't stop") In any case, these findings are important in view of proposals to remove the uncontrollability criterion from future versions of the DSM (e.g., Andrews et al., 2010). Our results, and those of Hallion and Ruscio (2013), suggest that uncontrollability is a dimension of GAD that uniquely explains individual differences in GAD and associated symptoms. Based on such findings, Hallion and Ruscio (2013) argued that the uncontrollability criterion should be retained in future versions of the DSM, due to its aid in the prediction of a wide range of clinical measures including clinical severity, comorbidity, and treatment-seeking, after controlling for excessiveness. Despite these recommendations, findings attesting to the differential predictive relevance of excessiveness and uncontrollability should be interpreted with caution given the considerable overlap in these dimensions of worry and the fact that the unique contribution of these features to the prediction of clinical outcomes is modest (i.e., medium to large effects for the full regression models, but small effects for the individual predictors of excessiveness and uncontrollability; cf. Table 2). It would be helpful to further explore the discriminant and incremental validity of these features in other samples (e.g., replication in other clinical samples, community-based samples) and other types of measurement (e.g., self-report measures, other interviews).

Several limitations should be considered when interpreting our results. First, the length of time separating assessments, while an improvement from a previous GAD reliability study (Gordon & Heimberg, 2011), may have impacted reliability due to changes in patient report or stressful life-events occurring between assessments. Second, the generalizability of our estimates of reliability should be noted due to the structured nature of training in ADIS-IV-L assessments at the Center for Anxiety and Related Disorders. We would expect reliability to be lower in clinical practice. Finally, with regard to moderator analyses, we had no choice but to impose a dichotomy on continuous variables (e.g., PSWQ score, days between assessments) because the method we used to test the

differential magnitude of ICCs requires a nominal independent variable (McGraw & Wong, 1996). Unfortunately, there is a loss of information about individual differences and loss of statistical power associated with imposing a dichotomy on continuous variables (MacCallum, Zhang, Preacher, & Rucker, 2002). In most instances, the cutoffs used were driven by the distribution of our sample data (e.g., PSWQ). Thus, the results may vary based on the cutoffs used.

Despite these limitations, the present study addresses the reliability and validity of the dimensional features of GAD, and reveals possibilities for future research in the area of classification. This study highlights the importance of dimensionally based assessment systems in order to evaluate severity and duration of disorders and their underlying constructs. Future studies should examine reliability and validity of GAD in a larger and more diverse community sample. Additionally, more research is needed to clarify the ongoing debates related to the excessiveness and uncontrollability criteria. Because our research was conducted with an adult sample, future research should investigate if estimates of reliability and validity of GAD features are similar in child and adolescent samples, as this research may aid in determining whether excessiveness, uncontrollability, or both should be retained in our diagnostic system. Finally, future studies should reexamine sources of diagnostic unreliability, particularly focusing on boundary issues between GAD and mood disorders. Because of the substantial overlap between anxiety and mood disorders and the deleterious effect of this overlap on reliability, our findings support the continued need for dimensional assessment in order to address high rates of comorbidity, boundary concerns, disorder severity, and loss of valuable clinical information. If dimensional symptom ratings are collected, researchers can continue to explore the shared features of anxiety and mood disorders and how these common dimensions may impact reliability, validity, assessment, and treatment (cf. Brown & Barlow, 2009).

References

- Abel, J. L., & Borkovec, T. D. (1995). Generalizability of *DSM-III-R* generalized anxiety disorders to proposed *DSM-IV* criteria and cross-validation of proposed changes. *Journal of Anxiety Disorders*, 9, 303–315. [http://dx.doi.org/10.1016/0887-6185\(95\)00011-c](http://dx.doi.org/10.1016/0887-6185(95)00011-c)
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Andrews, G., Hobbs, M. J., Borkovec, T. D., Beesdo, K., Craske, M. G., Heimberg, R. G., et al. (2010). Generalized worry disorder: a review of *DSM-IV* generalized anxiety disorder and options for *DSM-V*. *Depression and Anxiety*, 27, 134–147. <http://dx.doi.org/10.1002/da.20658>
- Beesdo, K., Winkel, S., Pine, D. S., Hoyer, J., Höfler, M., Lieb, R., et al. (2011). The diagnostic threshold of generalized anxiety disorder in the community: a developmental perspective. *Journal of Psychiatric Research*, 45, 962–972. <http://dx.doi.org/10.1016/j.jpsyres.2010.12.007>
- Borkovec, T. D., & Roemer, L. (1994). *Generalized anxiety disorder*. In: M. Hersen, & R. T. Ammerman (Eds.), *Handbook of prescriptive treatments for adults* (pp. 261–281). New York: Plenum Press.
- Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy*, 30, 33–37. [http://dx.doi.org/10.1016/0005-7967\(92\)90093-v](http://dx.doi.org/10.1016/0005-7967(92)90093-v)
- Brown, T. A., & Barlow, D. H. (1992). Comorbidity among anxiety disorders: implications for treatment and *DSM-IV*. *Journal of Consulting and Clinical Psychology*, 60, 835–844. <http://dx.doi.org/10.1037/0022-006x.60.6>
- Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on shared features of the *DSM-IV* anxiety and mood disorders: implications for assessment and treatment. *Psychological Assessment*, 21, 256–271. <http://dx.doi.org/10.1037/a0016608>
- Brown, T. A., Barlow, D. H., & Liebowitz, M. R. (1994). The empirical basis of generalized anxiety disorder. *American Journal of Psychiatry*, 151, 1272–1280.
- Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the *DSM-IV* anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, 107, 179–192. <http://dx.doi.org/10.1037/0021-843x.107.2.179>
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, 35, 79–89. [http://dx.doi.org/10.1016/S0005-7967\(96\)00068-X](http://dx.doi.org/10.1016/S0005-7967(96)00068-X)
- Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of *DSM-IV* anxiety and mood disorders: implications for the classification of emotional disorders. *Journal of Abnormal Psychology*, 110, 49–58. <http://dx.doi.org/10.1037/0021-843x.110.1.49>
- Brown, T. A., Marten, P. A., & Barlow, D. H. (1995). Discriminant validity of the symptoms constituting the *DSM-III-R* and *DSM-IV* associated symptom criterion of generalized anxiety disorder. *Journal of Anxiety Disorders*, 9, 317–328. [http://dx.doi.org/10.1016/0887-6185\(95\)00012-d](http://dx.doi.org/10.1016/0887-6185(95)00012-d)
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, 100, 316–336. <http://dx.doi.org/10.1037/0021-843x.100.3.316>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Di Nardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety Disorder Interview Schedule for DSM-IV: lifetime version (ADIS-IV-L)*. New York: Oxford University Press.
- Di Nardo, P. A., Moras, K., Barlow, D. H., Rapee, R. M., & Brown, T. A. (1993). Reliability of *DSM-III-R* anxiety disorder categories: using the Anxiety Disorders Interview Schedule-Revised (ADIS-R). *Archives of General Psychiatry*, 50, 251–256.
- Di Nardo, P. A., O'Brien, G. T., Barlow, D. H., Waddell, M. T., & Blanchard, E. B. (1983). Reliability of *DSM-III* anxiety disorder categories using a new structured interview. *Archives of General Psychiatry*, 40, 1070–1074.
- Gordon, D., & Heimberg, R. G. (2011). Reliability and validity of *DSM-IV* generalized anxiety disorder features. *Journal of Anxiety Disorders*, 25, 813–821. <http://dx.doi.org/10.1016/j.janxdis.2011.04.001>
- Hallion, L. S., & Ruscio, A. M. (2013). Should uncontrollable worry be excluded from GAD in *DSM-5*? A test of incremental validity. *Journal of Abnormal Psychology*, 122, 369–375. <http://dx.doi.org/10.1037/a0031731>
- Lovibond, P. F., & Lovibond, S. H. (1995). 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, 335–343. [http://dx.doi.org/10.1016/0005-7967\(94\)00075-U](http://dx.doi.org/10.1016/0005-7967(94)00075-U)
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40. <http://dx.doi.org/10.1037/1082-989X.7.1.19>
- Mannuzza, S., Fyer, A. J., Martin, L. Y., & Gallops, M. S. (1989). Reliability of anxiety assessment: I. Diagnostic agreement. *Archives of General Psychiatry*, 46, 1093–1101. <http://dx.doi.org/10.1001/archpsyc.1989.01810120035007>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. <http://dx.doi.org/10.1037/1082-989X.1.1.30>
- 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, 487–495. [http://dx.doi.org/10.1016/0005-7967\(90\)90135-6](http://dx.doi.org/10.1016/0005-7967(90)90135-6)
- Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., et al. (1992). The structured clinical interview for *DSM-III-R* (SCID): II. Multisite test-retest reliability. *Archives of General Psychiatry*, 49, 630–636.