Psychological Assessment
2011, Vol. 23, No. 4, 970–982

1040-3590/11/$12.00 DOI: 10.1037/a0024135

Assessing Psychological Flexibility: What Does It Add Above and Beyond Existing Constructs?

Andrew T. Gloster, Jens Klotsche, Samia Chaker, Katrin V. Hummel, and Jürgen Hoyer
Technische Universität Dresden

The construct of psychological flexibility (PF) is a central concept in acceptance and commitment therapy. It is defined as the process of contacting the present moment fully as a conscious human being and persisting in or changing behavior in the service of chosen values. PF is hypothesized to be an important aspect of healthy psychological functioning. Despite its potential importance, the distinctness of PF from other constructs has not been adequately demonstrated, and psychometric evaluations of measures designed to assess it are limited. This study aimed at extending current knowledge about PF by examining the construct in 2 help-seeking samples, including panic disorder with agoraphobia ($n = 368$), clinically relevant social phobia ($n = 209$), and 2 nonclinical samples including students ($n = 495$) and individuals visiting an employment office ($n = 95$). Results across all samples indicate that PF, as measured by the Acceptance and Action Questionnaire (2nd version; AAQ–II), is a unitary construct with a 1 factor model. PF correlated with other variables largely consistent with predictions, differentiated patients from healthy controls, and showed preliminary indications of treatment sensitivity. Incremental validity was partially demonstrated, especially for indices of functioning. Surprisingly, PF also explained unique variance above more established measures for some indices of symptomatology. Results suggest that PF adds some incremental clinical validity, yet further and more stringent tests are required to fully elucidate its strengths and limitations.

Keywords: psychological flexibility, AAQ–II, assessment, incremental validity, psychometrics

Psychological flexibility (PF) refers to the process of contacting the present moment and the thoughts and feelings it contains, without needless defense, fully as a conscious human being and, depending on what the situation affords, persisting or changing behavior in the service of chosen values (Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Hayes, Strosahl, & Wilson, 1999). PF is a broad, higher level construct used to capture several core, interconnected processes in one overarching term. The constituent processes implied in the definition stem from acceptance and commitment therapy (ACT); include experiential avoidance, acceptance, cognitive defusion, and mindfulness (see Hayes et al., 1999, 2006, for details); and are assumed to be relevant for the understanding of mental health, psychopathology and psychotherapy (Fletcher & Hayes, 2005). PF is not simply the absence of symptoms, however, nor is it synonymous with positive emotions. An individual with pervasive negative emotions can still interact with them flexibly, and one can demonstrate inflexibility despite positive emotions (Kashdan & Rottenberg, 2010; Schmaltz & Murrell, 2010).

The higher level of the construct and the fact that it is context dependent (i.e., a psychologically flexible response depends on the interaction between psychological content, the present moment, and chosen values) make the construct of PF increasingly utilized and, some authors argued, an important theoretical advance. At the same time, the higher level of the construct and its context dependency complicate its assessment in a self-report questionnaire format, as opposed to other options such as conducting a formal functional analysis (Fernandéz-Ballesteros, 2004; Iwata, Dorsey, Slifer, Bauman, & Richman, 1994).

Despite the apparent complications in its assessment, the construct of psychological flexibility is posited to be an important aspect of psychological health (Kashdan & Rottenberg, 2010). Indeed, the inverse, namely psychological inflexibility, has been associated with numerous indices of psychopathology (Hayes et al., 2006). Some research suggests that PF and its components function as a mediator of treatment change such that increases in psychological flexibility mediate or partially mediate changes in psychopathological symptoms such as depression (Zettle & Hayes, 1986; Zettle & Rains, 1989), anxiety (Forman, Herbert, Moitra, Yeomans, & Geller, 2007), pain (McCracken, Vowles, & Eccleston, 2005), nicotine addiction (Gifford et al., 2004), and psychosis (Bach & Hayes, 2002).

This article was published Online First July 18, 2011.

Andrew T. Gloster, Jens Klotsche, Samia Chaker, Katrin V. Hummel, and Jürgen Hoyer, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany.

The preparation of this article was supported in part by German Federal Ministry of Education and Research Grant BMBF (Project 01GV0615) and German Research Foundation Grant DFG HO 1900/5-1. The data from the panic disorder with agoraphobia (PD/AG) sample came from the German multicenter trial Mechanisms of Action in CBT (MAC). The MAC study was part of the BMBF Psychotherapy Research Funding Initiative. We thank Caroline Mey and Jens Siegert for their help in assessing the social phobia online data.

Correspondence concerning this article should be addressed to Andrew T. Gloster, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Chemnitzer Straβe 46, D-01187 Dresden, Germany. E-mail: gloster@psychologie.tu-dresden.de

Psychometric examinations of instruments designed to measure the construct remain limited. The Acceptance and Action Questionnaire (AAQ–I; Hayes et al., 2004), a nine or 16 item questionnaire (depending on the version), is the most widely used measure specifically designed to assess PF. A recent revision was undertaken with a main goal of generating new items that would yield better indices of internal consistency. This process resulted in the seven-item AAQ–II (Bond et al., in press). Both versions of the AAQ use terminology unspecific to any disorder so as to capture processes believed to apply universally. The AAQ has been found to correlate moderately with related constructs that aim to capture psychological processes (i.e., thought suppression, avoidance) and moderately to highly with numerous measures of psychopathology. Authors of the AAQ argued that the moderate to high inverse correlations observed with symptom measures have not been so high as to suggest identical constructs. Furthermore, these correlations were expected based on the premise that concurrent endorsement of symptoms are likely as the levels of this core psychological process decrease (i.e., psychological flexibility; Hayes et al., 2004).

Empirical examinations of the degree to which the construct of PF is unique from longer established constructs such as quality of life, psychological well-being, self-esteem, distress tolerance, or even some aspects of personality, such as neuroticism, remain scarce. Recent attention paid specifically to the construct of PF has brought researchers to rather cautious conclusions (Abramowitz, 2009; Kollman, Brown, & Barlow, 2009). Concerns with the construct are also indirectly derived from theoretical critiques of ACT—the therapeutic approach that explicitly targets psychological flexibility and whose proponents have developed the questionnaire designed to measure it (Bond et al., in press; Hayes et al., 2004). Briefly, these critiques suggest that ACT may represent a variation of cognitive behavioral therapy (CBT) but that the underlying mechanisms of action may not differ between the two (Arch & Craske, 2008; Hofmann & Asmundson, 2008). By way of extension, this suggests that targeting the construct of PF in therapy may lack sufficient differentiation from targeting other underlying therapeutic processes, such as altering attentional bias, reducing avoidance, and decreasing anxiety sensitivity—all of which are also targeted in other form of CBT. Indeed, two recent meta-analyses that failed to find substantial differences in efficacy between ACT and CBT also add to such concerns (Öst, 2008; Powers, Vörding, & Emmelkamp, 2009). Although the conclusions and findings of these studies have been challenged (Gaudiano, 2009; Levin & Hayes, 2009) and several studies suggest that ACT does work by a different mechanism from traditional CBT (see Hayes et al., 2006), questions about the validity of the construct and associated therapeutic targets remain.

Given these concerns, testing the borders of psychological flexibility in terms of conceptual benefit and incremental utility are important. A particularly striking recent result (Kollman et al., 2009) indicated that acceptance, one component of PF, was conceptually independent from overlapping constructs such as cognitive reappraisal and perceived emotional control. At the same time, acceptance was found to be practically unrelated to clinically relevant measures of worry, social interaction anxiety, environmental mastery, personal growth, purpose in life, and self-acceptance. Kollman et al. concluded that acceptance might lack clinical validity, although they also acknowledged that their assessment format limited their ability to detect such relations. Similarly, ambiguous results were found in a sample of students with high obsessive-compulsive (OC) symptomatology (Abramowitz et al., 2009). In this study, PF did not predict variance of OC symptoms above and beyond a measure of obsessive beliefs. The authors concluded that PF was less relevant than obsessive beliefs for the underlying processes that give rise to OC symptoms. This finding may be consistent with the theory underlying ACT and PF in that the presence of symptoms does not preclude PF. The authors go on to speculate, however, that the modification of beliefs is a more effective therapeutic target than PF. This suggests a conceptualization of PF that is partially dependent on symptoms. The high correlations between the AAQ and symptom measures suggest that this may be the case. As the authors acknowledge, however, this is currently unknown, may or may not be correct, and requires critical examinations of PF that control for concurrent symptomatology. Finally, recent experimental work found that although anxiety sensitivity and experiential avoidance (an aspect of PF) both related to self-reported fear in an emotional challenge procedure, anxiety sensitivity was more strongly related (Kelly & Forsyth, in press).

In light of the growing attention to PF, both positive and negative, the purpose of the current study was to examine PF and its uniqueness from other constructs. This was investigated in a two-step process in which we, first, examined the general psychometric properties of the AAQ–II and, second, tested specific hypotheses regarding the clinical utility of the AAQ–II. The initial step of presenting detailed psychometric data was necessary, given the paucity of published data on this measure to date. Data for our examinations came from four independent larger samples comprising both help-seeking individuals (i.e., individuals with panic disorder with agoraphobia and individuals with clinically relevant symptoms of social phobia) and predominately healthy participants (i.e., students and individuals visiting an employment center).

After psychometric analyses on factorial validity and reliability, we tested five specific hypotheses based on PF theory and recent empirical findings (Abramowitz et al., 2009; Bond et al., in press; Kashdan & Rottenberg, 2010; Kollman et al., 2009). First, we predicted that the AAQ–II would correlate positively with measures of functioning, quality of life, self-esteem, and the personality dimensions of Extraversion, Openness, Conscientiousness, and Negatively with all measures of psychopathology and the personality dimension of neuroticism. We expected the AAQ–II to be largely uncorrelated with age, sex, and the personality dimension of agreeableness. These predictions were made based on the posited role of PF as a key mediator of how symptoms are experienced. This, in turn, should translate to meaningful positive correlations with other human strengths. Second, because PF is a transdiagnostic construct, we hypothesized that PF would differentiate between nonclinical and help-seeking samples, but not between the two help-seeking groups. Third, given that the definition of PF includes the way individuals experience symptoms and that this is independent of a therapy approach, we predicted that PF would increase during CBT, even though CBT does not explicitly target PF as in ACT. Thus, we hypothesized a greater increase in PF relative to the waitlist (WL) and predicted that the increase of PF would occur more during the active phase of CBT (i.e., in situ exposure) relative to the beginning/preparatory phase.

Fourth, given that PF is conceptually contradictive to obsessive-compulsive disorder (OCD) symptomatology, we hypothesized that the AAQ–II would be negatively related to obsessions and compulsions but that PF would be more negatively correlated with compulsions than obsessions. This prediction is derived from PF theory, which suggests that occurrence of obsessions is largely beyond one's control, but if one responds to these unwanted cognitive events with PF (i.e., consistent with one's values) then it becomes possible to choose whether to respond to the obsessions with compulsions (Twohig et al., 2010). Finally, the definition of PF stresses pursuit of one's values even in the face of aversive conditions, and the ability to do so is considered an important feature of psychological health. Therefore, we hypothesized that when predicting functioning and impairment, PF would explain variance that could not be explained by other symptom measures or by the trait measures of anxiety sensitivity or neuroticism. This incremental contribution of PF was expected to be specific for the explanation of functioning and impairment but not for symptomatology itself.

Method

Participants

The present study included participants from four independent samples recruited from clinical and nonclinical settings in Germany. The first sample comprised patients reliably diagnosed with panic disorder with agoraphobia (PD/AG) as part of a large German multicenter randomized treatment outcome study ($n = 368$, 76.4% female, 23.6% male, 18 to 65 years, $M_{age} = 35.5$ years, $SD = 10.8$; Gloster et al., 2009, 2011). Patients were diagnosed with a standardized diagnostic interview (see below) administered by trained assessors. All patients in the PD/AG sample reported German to be their first language and indicated their nationality to be German. The mean number of reported years of education in this sample was 10.9 ($SD = 2.1$). Regarding living arrangements, 69.9% reported living with a partner, 18.9% reported alone. Current employment was reported by 59.1%, job training was reported by 26.6%, and unemployment was reported by 9.2%.

The second sample included help-seeking individuals with clinically relevant symptoms of social anxiety/social phobia (SP) recruited through four German Internet websites dealing with diagnosis and treatment of social anxiety disorder. Diagnostic status was assessed with an online format structured diagnostic interview (see below). Five hundred thirty-nine subjects clicked on the link and filled out the informed consent for participation. Of these, 224 (41.6%) quit the questionnaire after the first few questions and were excluded from further analysis. Due to a software failure, data from 93 participants (17.3%) were not completely saved. Participants who did not endorse any of the core social phobia criteria ($n = 13$) were eliminated for these analyses. The final sample comprised 209 help-seeking individuals with clinically relevant symptoms of social anxiety/social (69.4% female, 30.6% male, 16 years to 60 years, $M_{age} = 27.7$, $SD = 8.4$). All participants reported German as their first language and were Caucasian. Further, participants reported a mean of 12.7 years ($SD = 2.9$) of education.

The third sample comprised students (ST) at the Technische Universitaet Dresden attending courses of psychology, literature, and economics ($n = 495$). In order to determine test–retest reliability, 283 participants were tested once, and 212 were tested twice. The ST sample was predominately female (61.7%, 38.3% male), with a mean age of 22.5 years (18–46 years). Most students reported German as their first language (96.8%). All participants in the student sample reported at least 12 years of education.

The fourth sample comprised individuals in a waiting room of a municipal employment office (EO; $n = 95$). The employment office provides information regarding job vacancies, further education or training, or financial issues arising from unemployment. Individuals were predominantly female (51.6%, 48.4% male) with a mean age of 35.5 years (18–58 years, $SD = 13.5$). All participants were unemployed. Most participants had German nationality (93.7%) and reported a mean of 11.6 years ($SD = 3.3$) of education.

The AAQ total score could be calculated for $n = 368$ in PD/AG, $n = 495$ in ST, $n = 209$ in SP, and $n = 95$ in EO. Subjects with nonresponses to any of the AAQ items did not significantly differ from those with full response on any other measure. The combined sample (CS) thus consisted of 1,167 participants ($M_{age} = 28.1$, 16 to 65, $SD = 9.7$), a majority of whom were female (65.8%), German native speakers (98.1%), and Caucasian (99.5%). Prior to assessments, participants were informed about the study purposes, the voluntary nature of their participation, and data security, and participants gave informed consent. Data assessment procedures were approved by the University Research Ethics Committee.

Procedure

The PD/AG sample was part of a German multicenter therapy outcome study (Gloster et al., 2009, 2011). Diagnostic status was verified withy (a) a standardized computer-administered Composite International Diagnostic Interview (CIDI; Wittchen & Pfister, 1997; World Health Organization, 1990), (b) a score of ≥18 in the structured interview for the Hamilton Anxiety Scale (SIGH–A; Shear et al., 2001), and (c) a score of ≥4 on the Clinical Global Impression Scale (CGI; Guy, 1976). An extensive assessment battery was administered for the outcome study including those relevant for the present study.

The SP sample was recruited in a 4-week assessment period from websites dedicated to social anxiety disorder, two of which had an emphasis on blushing (Chaker, Hofmann, & Hoyer, 2010). All websites provided basic information on the diagnosis of social anxiety disorder, treatment options, personal experiences with treatment, and a self-help bulletin board. Within this framework, the study was described and participants logged on to a secure website to answer questionnaires. Participation was optional and unpaid. Symptomatology was assessed using an online format of the German version of the Anxiety Disorders Interview Schedule for the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR*; ADIS–IV; Brown, Di Nardo, & Barlow, 1994; Diagnostisches Interview bei Psychischen Störungen (DIPS for *DSM–IV–TR*); Schneider & Margraf, 2006) and following the online procedure used by Hammelstein and Soifer (2006). The online assessment procedure was programmed so that participants could only go to the next questionnaire after all items were filled out. Furthermore, it was possible to quit the study at any time. Participants who dropped out during the completion session were excluded.

Participants were categorized as having clinically relevant symptoms of social anxiety/phobia, if they endorsed one of the core symptoms of (a) having a marked and persistent fear of social or performance situations in which embarrassment may occur or (b) experiencing concerns or being afraid of others' judgments and also endorsing all the following symptoms: (a) experiencing at least moderate fear in one of the listed situations, (b) experiencing significant impairment, and (c) recognizing that the fear is excessive or unreasonable. Further questionnaires were also completed in the online format (described below).

Students at a German university were approached during lectures and were asked to complete a questionnaire packet. All students were informed about the goals of the study, the voluntary nature of their participation, and the data security before they gave oral consent.

Individuals visiting a German municipal employment office were approached while waiting for their appointments in the waiting room. Information about the goal of the study, the voluntary nature of participation, and the data security was given before the individuals signed the informed consent document. The assessments completed by each sample can be obtained from Table 3.

Measures

Anxiety Disorders Interview Schedule for *DSM–IV* (ADIS–IV; Brown et al., 1994). The ADIS–IV is a structured interview that assesses the diagnostic criteria of *DSM–IV* diagnoses. In the present study, the German version of the ADIS (DIPS for *DSM–IV–TR*; Schneider & Margraf, 2006) was administered in an online format. This version showed good retest-reliability scores, between .73 and .77, and a good interrater agreement of 88%, regarding the anxiety disorder sections.

Anxiety Sensitivity Index (ASI; Peterson & Reiss, 1993). The ASI is a screening instrument used to assess anticipatory fear or sensitivity to anxiety symptoms. It contains 16 items that are rated on a 5-point scale ranging from 0 (*very little*) to 5 (*very much*). Ehlers and Margraf (1993) reported that test scores demonstrated good internal consistencies for the German version across clinical and nonclinical samples. Patients diagnosed with panic disorder reported significantly higher scores on the ASI than did patients diagnosed with other anxiety disorders (Apfeldorf, Shear, Leon, & Portera, 1994).

Beck Depression Inventory (2nd ed.; BDI–II; Beck, Steer, & Brown, 1996). The second edition of the Beck Depression Inventory is a 21-item self-report questionnaire designed to measure depression symptoms according to *DSM–IV* criteria. Participants rate symptom severity, in 19 items on a 4-point Likert scale ranging from 0 to 3, with respect to the past 2 weeks. Further, two items allow for an indication of an increase or a decrease in sleep or appetite. The German version used in the present study has shown strong reliability and validity of BDI–II scores in clinical and nonclinical samples (Hautzinger, Keller, & Kühner, 2006).

Big Five—16 Adjective Measure (BF–16 AM; Herzberg & Brähler, 2006). The BF–16 AM is an instrument designed to assess five personality traits, usually referred to as The Big Five (i.e., Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness). The BF–16 AM refers to the lexical approach to personality description and assesses the Big Five via 16 adjectives, which were rated on a 7-point scale ranging from 1 (*disagree strongly*) to 7 (*agree strongly*). Validation of the BF–16 AM revealed reasonable internal consistencies for the scale scores ranging from $\alpha = .57$ (openness) to $\alpha = .74$ (conscientiousness) in a nonclinical sample. Correlations of the BF-16 AM with the NEO–Five Factor Inventory (NEO-FFI) ranging from .64 (agreeableness) to .92 (neuroticism) indicate satisfying convergence (Herzberg & Brähler, 2006).

Brief Symptom Inventory (BSI; Derogatis, 1993). The BSI (the self-report form of the Symptom Checklist 90-R) assesses severity of somatic and psychological symptoms on nine subscales over the past week. It contains 53 items that are scored on a 5-point scale ranging from 0 to 4 with increasing severity of symptoms. The German version (Franke, 2000) used in the present study demonstrated strong reliability for scale scores and validity of score interpretations in a sample of psychotherapeutic outpatients (Geisheim et al., 2002).

Clinical Global Impression (CGI; Guy, 1976). The CGI is a brief observer-rated three-item scale that measures illness severity, global improvement or change, and therapeutic response. The CGI is rated on a 7-point scale, and the German version was employed. Zaider, Heimberg, Fresco, Schneier, and Liebowitz (2003) supported the utility of the CGI revealing positive correlations of the CGI with self-report and clinician-administered measures across different psychological domains.

Composite International Diagnostic Interview (CIDI–M/DIA–X; Wittchen & Pfister, 1997). Diagnostic status for the PD/AG sample was established using the CIDI–M/DIA–X. The DIA–X/M–CIDI is a standardized clinical interview that assesses symptoms, syndromes, and diagnoses of mental disorders according to *DSM–IV* and *International Classification of Diseases* (10th ed.; *ICD-10*). Additional information is obtained about onset, duration, and severity. The administered computerized version revealed good reliability for core *DSM–IV* symptom questions and good diagnostic agreement (Wittchen & Pfister, 1997), and this version is usually well accepted by respondents (Hoyer, Ruhl, Scholz, & Wittchen, 2006).

Depression Anxiety Stress Scales—21-Item version (DASS–21; Antony, Bieling, Cox, Enns, & Swinson, 1998; Lovibond & Lovibond, 1995). The DASS-21 measures the core symptoms of depression, anxiety, and stress over the past week on three subscales. Respondents are asked to rate how much each statement applied to them on a 4-point Likert scale. A German version of the DASS-21 (Köppe, 2001) was used in the present study, retaining the 1-week timeframe of the English version. Psychometric data on the German version showed Cronbach's α ranging from .77 to .88 for scores of the three subscales and .92 for the total score (Köppe, 2001).

EuroQOL 5D (EQ-5D; EuroQol Group, 1990). The EQ-5D is a standardized short measure of health related quality of life. It contains five items and a visual analog scale (VAS) asking patients to rate their health on a scale of 0–100. The EQ-5D has been used in a German version as provided by the EuroQol Group. Scores of the EQ-5D showed moderate levels of test–retest reliability in a multinational study and have been validly interpreted in relation to known groups with health and emotional problems (Ravens-

Sieberer et al., 2010; von der Schulenburg, Claes, Greiner, & Uber, 1998).

Hamilton Anxiety Scale—Structured Interview Guide (SIGH–A; Hamilton, 1959; Shear et al., 2001). The SIGH–A is a 14-item instrument filled out by the clinician interviewing the patient. It consists of two groups of items measuring the severity of physical and psychological symptoms. Each item is rated on a 5-point scale. The SIGH–A can be administered evaluate the therapeutic process. Shear et al. (2001) reported high interrater reliability for ratings (.89–.99) and internal consistencies (.79–.88) for rating scores in patients with and without current anxiety diagnoses.

Obsessive-Compulsive Inventory—Revised (OCI–R; Foa et al., 2002). The OCI–R is a revised 18-item short form of the original Obsessive-Compulsive Inventory (OCI; Foa, Kozak, Salkovskis, Coles, & Amir, 1998). The OCI–R assesses the distress associated with obsessions and compulsions over the last week on six subscales. Symptoms are rated on a 5-point-Likert scale. The German version of the OCI–R revealed a Cronbach's α of .85 for the total scale score and between .76 and .95 for the subscale scores (Gönner, Leonhart, & Ecker, 2008).

Panic and Agoraphobia Scale (PAS; Bandelow, 1995). The PAS assesses severity in panic disorder with/without agoraphobia in a self-report format and is sensitive to treatment change. The scale consists of five subscales with a total of 13 items rated on a 5-point scale. PAS scores in a sample of patients with panic disorder showed satisfactory levels of reliability and validity (Bandelow, 1995).

Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965). The RSE consists of 10 items measuring global self-esteem with items rated on a 4-point scale. The German version revealed good internal consistencies for the RSE scores ranging from .84 to .85 (von Collani & Herzberg, 2003).

Retrospective Self-Report of Inhibition (RSRI; Reznick, Hegeman, Kaufmann, Woods, & Jacobs, 1992). The RSRI is a 30-item retrospective questionnaire measure of behavioral inhibition including two subscales: Social/School and Fear/Illness. The present study used the German version of the RSRI (Rohrbacher et al., 2008). Moderate to good validity of scale scores for a representative community sample was demonstrated with internal consistencies ranging from .69 to .74.

Self-Consciousness Scale (SCS; Fenigstein, Scheier, & Buss, 1975). The questionnaire for the assessment of dispositional self-awareness (SAM-questionnaire; Filipp & Freudenberg, 1989) is a 27-item German adaptation of the original SCS. It assesses self-consciousness on two scales (private and public). The SAM questionnaire was validated in several studies with internal consistencies for scores ranging from .71 to .88 and with test–retest reliability for scores ranging between .72 and .84. Further analysis revealed good convergent and discriminant validity (Filipp & Freudenberg, 1989).

Intervention

The sensitivity to detect treatment change was examined in the PD/AG sample. Patients in this group underwent a course of 12 manualized sessions of CBT for PD/AG administered twice weekly (Gloster et al., 2009; Gloster et al., 2011). The CBT consisted of psychoeducation, functional analysis, rationale for exposure, interoceptive exposure, in situ exposure, discussion of anticipatory anxiety, and relapse prevention but did not explicitly target psychological flexibility. Assessments occurred at baseline, at an intermediate point in treatment (i.e., following psychoeducation and interoceptive exposure—after the 4th session), at posttreatment, and at a 6 month follow up.

Statistical Analyses

All analyses reported in this article were conducted on the seven-item version of the AAQ–II. Confirmatory factor analysis (CFA) was used to test whether the AAQ–II factor structure was one-dimensional, as reported elsewhere (Bond et al., in press). An unrestricted one-factor model assuming no residual covariance among the AAQ–II items was fit and examined for sources of possible model misspecification, as indicated by the modification index. Research has shown that it is frequently necessary to assume correlated measurement errors among items with similar content to avoid method effects, particularly with items using the same key terms (Brown, 2003; Floyd & Widaman, 1995; Gerbing & Anderson, 1984). Thus, the final model was specified by assuming correlated measurement errors between Items 1 and 4 ("My painful experiences and memories make it difficult for me to live a life that I would value" and "My painful memories prevent me from having a fulfilling life") and between Items 2 and 3 ("I am afraid of my feelings" and "I worry about not being able to control my worries and feelings"). Acceptable model fit was assessed with standardized root-mean-square residual (SRMR; $\leq$0.08) and comparative fit index—Tucker Lewis index (CFI–TLI) $\geq$0.9), applying the guidelines for these fit indices suggested by Hu and Bentler (1998).

A multiple-group CFA was also conducted in order to test the stability, generalizability, and population homogeneity of the one factor solution. This procedure consists of multiple steps as recommended by Brown (2006). The one factor model is tested for each subsample separately at first. The acceptable model fit in each subsample provides the base for further steps. In the second step, the hypotheses of an identical factor structure is tested simultaneously across the four subsamples. The equality of factor loading is tested in the last step by constraining the factor loadings to be invariant across samples. If the factor loadings are invariant, it indicates a similar correlation structure across samples. CFA and multiple-group CFA were conducted in MPLUS 5.1 (Muthén & Muthén, 2008), applying the weighted least squares estimator based on a diagonal weight matrix.

In order to facilitate comparison across independent studies, comprehensive psychometric indices were calculated. Internal consistency was estimated with the formula discussed by Raykov (2001) within the hypothesized one factor CFA model assuming correlated measurement errors. The CFA-based approach for calculating scale reliability allows for taking the specified error structure into account, in contrast to Cronbach's alpha. Test–retest reliability was calculated for the PD/AG-waitlist patients ($N = 63$) and the ST sample with Pearson correlations. Neither sample received treatment during the test–retest interval.

Hypothesis 1 (construct validity) was examined with Pearson correlation coefficients. Hypothesis 2 (differentiation between samples of clinically relevant help-seeking and normative control samples) was tested with analysis of variance. Three planned

contrasts were calculated: all nonclinical controls versus all help-seeking individuals; PD vs. SP and students versus EO. Hypothesis 3 (sensitivity to treatment) was examined by calculating a reliable change index (RCI) for each participant. The RCI represents change in a patient's individual score standardized with the standard error for the difference for the retest period used. Those participants with a RCI greater than 1.96 were categorized as having achieved reliable improvement. For these analyses, the reliability estimates were obtained from the test–retest coefficients. Hypothesis 4 (stronger correlation with compulsions than obsessions) was examined by calculating differences between correlations within a sample (i.e., correlated correlations), then tested using the Fisher *z* transformation (Meng et al., 1992). For Hypothesis 5 (incremental validity), multiple regression analyses were conducted to test the additional predictive value of the AAQ–II for outcome variables (i.e., anxiety, panic symptoms, agoraphobic avoidance, overall functioning, and various indices of functional impairment) above and beyond the predictive value of established constructs (i.e., anxiety sensitivity, symptoms of depression, anxiety, and stress as well as neuroticism). Likelihood ratio tests were performed to test for the additional predictive value of the AAQ besides the other predictor variables.

Results[1]

Psychometric Properties

Factor structure. The one-factor model for the AAQ–II tested in the total sample did not fit the data well (*SRMR* = .049; root-mean-square error of approximation [*RMSEA*] = .195; *CFI* = .88; *TLI* = .82). Localized area of strain in that solution suggested correlated residuals between Items 2 and 5 (modification indices = 442.3) and between Items 3 and 4 (modification indices = 162.4). The model with correlated residuals yielded a significantly improved model fit, $\Delta\chi^2(2) = 532.27$, $p < .001$, with an acceptable overall model fit (*SRMR* = .024; *RMSEA* = .059; *CFI* = .98; *TLI* = .97). The pattern of unacceptable model fit for the one factor solution and an improved, acceptable fit for the model with correlated residuals for the item pairs (i.e., two–five and three–four) was replicated for samples PD/AG, SP, ST, and EO, as presented in Table 1.

Table 1
Model Fit of the One Factor Model of the AAQ–II Assuming Correlated Residuals for the Item Pairs Two–Five and Three–Four

Sample	*N*	*SRMR*	*RMSEA*	*CFI*	*TLI*
Total	1,167	.024	.059	.98	.97
PD/AG	368	.036	.061	.96	.93
SP	209	.022	.058	.99	.99
ST	495	.032	.067	.98	.96
EO	95	.049	.055	.93	.90

Note. AAQ–II = Acceptance and Action Questionnaire (2nd version); PD/AG = panic disorder with agoraphobia; SP = clinically relevant symptoms of social anxiety/social phobia; ST = student sample; EO = employment office sample; *SRMR* = standardized root-mean-square residual; *RMSEA* = root-mean-square error of approximation; *CFI* = comparative fit index; *TLI* = Tucker Lewis index.

As a final step, a multiple-group CFA was conducted. The one-factor model with correlated measurement errors between Items 2 and 5 and Items 3 and 4 resulted in an acceptable model in all four subsamples, as reported above. The multiple group CFA for the four samples was calculated to assess simultaneously an identical factor structure. This analyses yielded an acceptable model fit, $\chi^2(114) = 275.53$; $p < .001$; *SRMR* = .053; *RMSEA* = .065; *CFI* = .97; *TLI* = .97, indicating a one factor solution across the four samples. Next, the equivalence of factor loadings between Samples 1, 2, 3, and 4 was tested. The model assuming equal factor loadings and intercepts has an overall acceptable fit, $\chi^2(132) = 359.38$; $p < .001$; SRMR = .076; *RMSEA* = .069; *CFI* = .94; *TLI* = .94, suggesting the AAQ–II has the same meaning and structure for the four subsamples (Brown, 2006).

Internal consistency. Reliability of AAQ–II scores based on the Raykov estimation (2001) for the total sample was .93, thus exceeding the alpha value of the nine-item AAQ, which was .70 (Hayes et al., 2004). Across the four samples, the internal consistency ranged from good to excellent: .84 for the SP sample and EO, .92 for the PD/AG patients, and .97 for the ST sample.

Test–retest reliability. As predicted, the reliability of AAQ–II scores in both samples was good, as can be seen in Table 2. Patients in the wait list of the PD/AG sample (i.e., they did not receive therapy during this period) completed the AAQ–II three times, resulting in two separate retest periods. The retest validity was good across both periods, suggesting stability even across multiple testings.

Tests of Construct and Incremental Validity

Hypothesis 1: Zero-order correlations with other constructs. Pearson correlation coefficients were calculated between the AAQ–II and other measurement instruments used in this study (see Table 3). All observed correlations were in directions consistent with theoretical assumptions. For instance, the AAQ–II was inversely and predominantly moderately correlated with constructs of depression, anxiety, stress, anxiety sensitivity, behavioral inhibition, and neuroticism and was positively and moderately to strongly correlated with constructs such as self-esteem and extraversion. The magnitude of correlations was negligible for the number of panic attacks experienced, agreeableness, sex, and age (in three of four samples).

Patterns of correlations across subscales of an instrument can be especially informing because method variance is held constant, thereby facilitating comparisons of the facets measured by the subscales. With respect to the Hamilton Anxiety Scale, items assessing psychic symptoms were more strongly related than were physical symptoms (−.30 vs. −.13). Patterns across the well-established constructs of the Big Five personality traits are also noteworthy and consistent with a recent review on the relation between the Big Five and PF (Kashdan & Rottenberg, 2010). Neuroticism correlated most strongly and represented the lone inverse relationship. The correlation with extraversion was posi-

[1] All results reported in this article reflect the seven-item AAQ-II. Analyses were repeated using the 10-item version that includes three additional reverse scored items that were removed from the final version. All results were nearly identical and are available upon request.

Table 2
Test–Retest Reliability for the AAQ–II

Sample	Interval in weeks	*N*	Test		Retest		Retest–reliability
			M	*SD*	*M*	*SD*	
ST	2.0	212	40.0	7.2	41.0	7.6	.85
PD/AG–WL: Baseline—Intermediate	3.8	63	31.8	8.3	33.9	9.8	.75
PD/AG–WL: Intermediate—Posttreatment	5.1	63	31.8	8.3	33.8	9.5	.74

Note. AAQ–II = Acceptance and Action Questionnaire (2nd version); ST = student sample; PD/AG–WL = panic disorder with agoraphobia sample—waitlist patients only.

tively and strongly related, whereas conscientiousness and openness were moderately related. Consistent with theory, only agreeableness did not correlate meaningfully with psychological flexibility.

Hypothesis 2: Differentiation between nonclinical and help-seeking samples (concurrent validity). Consistent with our first hypothesis, PF was significantly higher in the two nonpatient groups than in the two help-seeking groups, 40.4 (7.6) versus 31.3 (9.8), $F(1, 1{,}171) = 311.50$, $p < .001$. Also consistent with our hypothesis, the mean AAQ–II score did not differ between the ST and EO groups, 40.4 (7.5) versus. 40.2 (8.0), $F(1, 595) = 0.07$, $p = .791$. Contrary to our hypotheses, however, the PD/AG and SP groups were significantly different, 33.3 (10.0) versus 27.9 (9.2), $F(1, 574) = 4.8$, $p < .001$.

Hypothesis 3: Treatment sensitivity. In order to examine the sensitivity of PF to treatment, the change in AAQ–II scores observed between baseline and posttreatment were compared for the active patients and WL of the PD/AG sample. As expected, patients receiving treatment reported significantly larger improvement in PF (5.6, $SD = 8.0$, $d = 0.63$) than in the WL ($-.14$, $SD = 6.8$, $d = 0.01$; mean difference in change = 5.7), $F(318, 1) = 27.9$, $p < .001$, $d = 0.72$. The degree of change was also examined for clinical significance. The proportion of PD/AG patients in active treatment who obtained reliable improvement was significantly lower for the interval between baseline and intermediate assessment than the interval between intermediate and posttreatment (3.5% vs. 13.8%), $\chi^2(1) = 25.56$, $p < .001$.

Hypothesis 4: Differentiation between compulsions and obsessions. With respect to OCD, the magnitude of correlations with subscales reflecting different symptom types were not uniform. Contrary to our hypothesis, however, obsessing correlated higher with PF ($r = -.60$) than all compulsion subscales (washing: $r = -.28$, $z = 9.6$, $p < .001$; hoarding: $r = -.30$, $z = 9.6$, $p < .001$; checking: $r = -.27$, $z = 10.0$, $p < .001$; neutralizing: $r = -.28$, $z = 9.4$, $p < .001$).

Hypothesis 5: Prediction of unique variance above established constructs (incremental utility). Hierarchical multiple regression analyses were conducted to test the hypothesis that PF incrementally predicts functioning and avoidance but not symptomatology (i.e., panic symptoms). These analyses were conducted only for the two help-seeking samples (see Table 4). In the PD/AG sample, the AAQ–II explained unique variance above and beyond that explained by the ASI and BDI–II for overarching anxiety symptomatology (i.e., HAM-A; contrary to the hypothesis) and patient functioning (consistent with the hypothesis), but not for patients' panic symptoms (consistent with the hypothesis) or agoraphobic avoidance (contrary to the hypothesis). In the SP sample, the AAQ–II added explained unique variance above and beyond that explained by the three subscales of the DASS-21 and neuroticism for impairments in daily life, free time, and social contacts—all consistent with our hypotheses.

Discussion

In this study, we examined PF and its uniqueness from other constructs. Generally, the data indicate that the AAQ–II, a measure of PF, has sound psychometric properties and adds to the explanation of clinically relevant variables, such as functioning and impairment, beyond well-validated measures of depression, anxiety, and stress, as well as anxiety sensitivity and neuroticism. It therefore seems as if the underlying construct, PF, helps elucidate some aspects of mental health, namely impairment and functioning, which would otherwise remain unexplained. The construct PF thus makes a unique and incremental contribution, though not uniformly consistent with a priori predictions. Although caution is still advised when referring to this or any other higher level construct, the findings in this study suggest that psychological flexibility is worthy of further consideration and close scrutiny in research targeting determinants of psychopathology and treatment mechanisms.

To our knowledge, the psychometric properties reported in this study are some of the first reported beyond the original report (Bond et al., in press) and thereby represent an important independent test. Test scores on the AAQ–II yielded good to excellent indices of reliability across four diverse samples. A one-factor solution with correlated residuals was found for the AAQ–II across two help-seeking samples for anxiety disorders (i.e., panic disorder/agoraphobia and individuals with clinically relevant symptoms of Social Phobia) and the two normative samples (i.e., students and people visiting an employment office). Thus, as measured by the AAQ–II, PF can be considered a unitary construct, and this information is based on a large data set ($N > 1{,}000$). The correlated residuals result from parallel wording in two-item pairs. This finding is consistent with the original version (Bond et al., in press). The unitary factor structure emerged despite the fact that the items of the AAQ–II were developed to assess the six core processes/facets targeted in ACT. Taken together with the good to excellent indices of internal consistency, the results suggest that the AAQ–II does not simply measure one or more of the core processes to the exclusion of others, rather, the totality of the items reflect the higher level construct that can be called psychological flexibility.

Table 3
Correlations Between AAQ–II and Other Measures

Construct	Measure	Sample	*r* with AAQ–II
	Symptomatology		
Depression	BDI–II	ST	−.73
		PD/AG	−.51
		EO	−.54
	DASS-D	SP	−.70
		ST	−.66
Anxiety	DASS-A	SP	−.59
		ST	−.53
Stress	DASS-S	SP	−.57
		ST	−.50
Severity of panic symptoms	PAS total	PD/AG	−.16
	PAS number of panic attacks	PD/AG	−.08
	PAS agoraphobia avoidance	PD/AG	−.10
	PAS anticipatory anxiety	PD/AG	−.01
	PAS disability	PD/AG	−.11
	PAS worries	PD/AG	−.19
Severity of anxious symptomatology	HAM-A total	PD/AG	−.29
	HAM-A psych	PD/AG	−.30
	HAM-A physic	PD/AG	−.13
Obsessive-compulsive	OCI total	ST	−.49
	OCI washing	ST	−.29
	OCI obsessing	ST	−.62
	OCI hoarding	ST	−.31
	OCI ordering	ST	−.29
	OCI checking	ST	−.28
	OCI neutralizing	ST	−.28
Subjective impairment	BSI total	EO	−.72
	Somatization	EO	−.46
	Obsession-compulsion	EO	−.62
	Interpersonal sensitivity	EO	−.69
	Depression	EO	−.55
	Anxiety	EO	−.61
	Hostility	EO	−.58
	Phobic anxiety	EO	−.59
	Paranoid ideation	EO	−.55
	Psychoticism	EO	−.55
	Nonspecific traits		
Self-esteem	RSE	SP	.67
Behavioral inhibition	RSRI total	SP	−.58
	RSRI fear/illness	SP	−.46
	RSRI social/school	SP	−.53
Self-consciousness	SCS public	SP	−.50
	SCS private	SP	−.35
	SCS total	SP	.67
Anxiety sensitivity	ASI	PD/AG	−.52
	Personality traits		
Neuroticism	BF-16-AM-N	SP	−.63
Extraversion	BF-16-AM-E	SP	.49
Openness	BF-16-AM-O	SP	.18
Agreeableness	BF-16-AM–A	SP	.07
Conscientiousness	BF-16-AM–C	SP	.35
	Global functioning		
Total functioning	CGI	PD/AG	−.12
Quality of life	EQ-5D	PD/AG	.34
	Nonpsychological variables		
Age	Years	PD/AG	.12
	Years	SP	.11^{ns}
	Years	ST	−.01^{ns}
	Years	EO	.07^{ns}
Sex	Male or female	PD/AG	−.00^{ns}
	Male or female	SP	.01^{ns}
	Male or female	ST	−.00^{ns}
	Male or female	EO	−.12^{ns}

Note. All correlations are significant unless indicated by *ns*. AAQ–II = Acceptance and Action Questionnaire (2nd version); ST = student sample; PD/AG = panic disorder with agoraphobia sample; SP = clinically relevant symptoms of social anxiety/social phobia; EO = employment office sample; BDI–II = Beck Depression Inventory (2nd ed.); DASS-D, DASS-A, and DASS-S = Depression Anxiety Stress Scales 21; PAS = Panic and Agoraphobia Scale; HAM-A = Hamilton Anxiety Scale; OCI = Obsessive-Compulsive Inventory-Revised; BSI = Brief Symptom Inventory; RSE = Rosenberg Self-Esteem Scale; RSRI = Retrospective Self-Report of Inhibition; SCS = Self-Consciousness Scale; ASI = Anxiety Sensitivity Index; BF-16-AM-N = Big Five–16 Adjective Measure–Neuroticism; BF-16-AM-E = Big Five–16 Adjective Measure–Extraversion; BF-16-AM-O = Big Five–16 Adjective Measure–Openness; BF-16-AM-A = Big Five–16 Adjective Measure–Agreeableness; BF-16-AM-C = Big Five–16 Adjective Measure–Conscientiousness; CGI = Clinical Global Impression; EQ-5D = EuroQOL 5D; Psych = psychic anxiety; Physic = somatic anxiety.

An extensive sampling of instruments across multiple studies is crucial to properly understand a new construct's unique boundaries and overlapping areas with more established constructs. As measured in the four diverse samples here, the pattern of scores with other constructs were largely consistent with the theoretical basis of PF. PF was positively and strongly correlated with the construct of self-esteem, yet not so highly that its differentiation should be questioned. Likewise, PF, although highly correlated with these variables, does not seem to be simply the same as high functioning or quality of life or extraversion. PF was negatively correlated with all measures of global psychopathology—most strongly with those of depression—and the personality trait of neuroticism. Equally important, PF does not seem to be related to the construct of agreeableness or to nonpsychological variables such as sex and age. Interestingly, whereas the pattern of correlations between the AAQ–II and Big Five personality traits is consistent with the original publication and a recent review (Bond et al., in press; Kashdan & Rottenberg, 2010), the magnitude of correlations was higher in the present sample except for Agreeableness. It is currently unclear whether this results from the sample utilized (individuals with clinically relevant symptoms of social anxiety/social phobia), the different measurements of personality (NEO–FFI vs. BF-16-AM), other factors, or the combination of these.

This study is one of the first to document the test–retest reliability over varying time frames, samples, and frequency of assessment. Test–retest analyses suggest that PF is a stable construct in nontreated individuals. The higher test–retest reliability observed in the ST sample relative to the WL patients of the PD/AG sample may reflect a slightly lower stability in clinical cases, at least while waiting for treatment. If replicated, this observation may point to a differential stability of PF as a function of absolute

Table 4
Hierarchical Multiple Regression Analyses of Psychological Flexibility and Baseline Severity

Sample	Predictor	Measure	Model 1 β	Model 1 *p*	Model 2 β	Model 2 *p*	Δ in R^2(*p*)
			HAM-A total				
PD/AG	Anxiety sensitivity	ASI	.21	.000	.18	.000	
	Depressive symptomatology	BDI–II	.43	.000	.40	.000	
	Psychological flexibility	AAQ–II			−.09	.047	.022 (.038)
			Panic symptoms				
PD/AG	Anxiety sensitivity	ASI	.21	.000	.20	.001	
	Depressive symptomatology	BDI–II	.13	.007	.11	.068	
	Psychological flexibility	AAQ–II			−.05	.104	.012 (.191)
			Agoraphobic avoidance				
PD/AG	Anxiety sensitivity	ASI	−.01	.907	−.02	.608	
	Depressive symptomatology	BDI–II	.20	.000	.17	.002	
	Psychological flexibility	AAQ–II			−.06	.276	.019 (.128)
			Functioning				
PD/AG	Anxiety sensitivity	ASI	.12	.020	.08	.135	
	Depressive symptomatology	BDI–II	.20	.000	.17	.004	
	Psychological flexibility	AAQ–II			−.10	.039	.024 (.012)
			Impairment in daily life				
SP	Depressive symptomatology	DASS-D	.34	.001	.21	.007	
	Anxious symptomatology	DASS-A	.43	.001	.38	.001	
	Stress symptomatology	DASS-S	−.24	.001	−.26	.001	
	Neuroticism	BF-16 AM-N	.23	.001	.13	.043	
	Psychological flexibility	AAQ–II			−.31	.001	.060 (.001)
			Impairment in free time				
SP	Depressive symptomatology	DASS-D	.51	.001	.38	.001	
	Anxious symptomatology	DASS-A	.19	.011	.15	.044	
	Stress symptomatology	DASS-S	−.17	.028	−.19	.013	
	Neuroticism	BF-16 AM-N	.20	.004	.10	.139	
	Psychological flexibility	AAQ–II			−.31	.001	.043 (.001)
			Impairment in social contacts				
SP	Depressive symptomatology	DASS-D	.46	.001	.25	.001	
	Anxious symptomatology	DASS-A	.19	.009	.10	.125	
	Stress symptomatology	DASS-S	−.19	.016	−.21	.004	
	Neuroticism	BF-16 AM-N	.26	.001	.10	.090	
	Psychological flexibility	AAQ–II			−.52	.001	.108 (.095)

Note. Panic symptoms, agoraphobic avoidance, and functioning were measured with the Clinical Global Interview (CGI). Impairment in daily life, Impairment in free time, and Impairment in social contacts were measured with the Anxiety Disorders Interview Schedule for *DSM–IV*. HAM-A = Hamilton Anxiety Scale; ASI = Anxiety Sensitivity Index; BDI–II = Beck Depression Inventory (2nd ed.); AAQ–II = Acceptance and Action Questionnaire (2nd version); DASS-D, DASS-A, and DASS-S = Depression Anxiety Stress Scales 21; BF-16 AM-N = Big Five–16 Adjective Measure—Neuroticism Subscale; PD/AG = panic disorder with agoraphobia; SP = clinically relevant symptoms of social anxiety/phobia; *DSM–IV* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.).

level of PF. That is, high and low levels of PF may be more likely to be stable—be it in a linear or curvilinear relation—perhaps in response to the contextual effects associated with such states. This remains highly speculative, however, until replications and controlled studies are conducted.

It has been postulated that the absence of PF contributes to psychopathology, though not to any specific diagnostic condition (Bond et al., in press). We therefore tested how well the AAQ–II could differentiate between clinically relevant help-seeking individuals and nonclinical samples. The observed difference in PF between these samples was a critical, albeit limited, test of PF's validity. The higher level of PF in the PD/AG sample compared with the SP was somewhat surprising. It could be argued that the observed 5 point difference is of minor importance and is statistically significant simply as a result of the large sample size or an artifact of the different response format/research setting (in a clinic vs. online). Alternately, the lower level of PF reported by the SP sample may represent an important psychological deficit that might be specific for individuals with social phobia who suffer primarily from fear of blushing (Chaker, Hofmann, & Hoyer, 2010). These individuals may attempt to suppress a response that is natural and cannot be altered (blushing) due to their striving to

react perfectly in social situations. Their efforts to control this objectively harmless and uncontrollable response can be considered antagonistic to PF. Furthermore, these individuals may have reported lower levels of PF because stressors occur more often and across a more extensive array of situations, precisely because social situations are more difficult to avoid than agoraphobic situations. Indeed, the interaction between social anxiety and emotion suppression has been associated with a lower number of positive events (Kashdan & Steger, 2006). Research is needed to test the degree to which PF differentially applies across diagnoses and how this interacts with the degree of functional impairment across multiple domains of life (e.g., social, work, health, etc.).

Analyses examining sensitivity to treatment showed that active PD/AG patients in a treatment that did not actively target PF (CBT) nevertheless resulted in a greater increase in PF compared with the WL. Nevertheless, the overall proportion of PD/AG with clinically significant change as measured by the RCI was low. As the PD/AG patients underwent a treatment that did not explicitly target PF, the pattern of change across treatment may be more informative for treatment sensitivity. Of those patients who did achieve reliable improvement in PF, nearly all of them occurred between the intermediate assessment and postassessment. That is, the change occurred during the core phase of the treatment (when patients confront situations that have long since been avoided during exposure in situ) and not during the initial phase (when information on anxiety is given and motivation for exposure treatment is enhanced). This is consistent with theory to the extent that exposure promotes PF. Although we are unaware of any clinical studies that directly address this question, ACT manuals for anxiety utilize exposure-like procedures (Eifert & Forsyth, 2005). These results provide some support for the treatment sensitivity in terms of statistically significant change and to a lesser degree in terms of clinically significant change as observed in a non-ACT treatment. The reasons for the overall low proportion of patients who demonstrated reliable improvement is currently unclear but is likely due to a combination of the treatment, the sample, or the relatively brief therapy (12 sessions in, on average, 8 weeks).

Obsessive-compulsive (OC) symptoms may be particularly informative about the boundary conditions of PF. By definition, obsessions are unwanted psychological experiences, and compulsions can be seen as a rigid response repertoire to the obsessions. As such, obsessive-compulsive symptoms, in particular compulsions, are not met with a stance of PF. This study found that PF is differentially related to subcomponents of OC symptoms, but in the opposite direction of our a priori predictions. At the zero-order level, PF was strongly and inversely correlated with obsessions (−.62), whereas compulsive symptoms (i.e., washing, hoarding, ordering, checking, and neutralizing) correlated less strongly (−.28 to −.31). The pattern was in contrast to correlations observed in a selected group of high-OC students (Abramowitz, 2009), where all correlations were lower, and the OCI subscale measuring compulsions (washing) was highest at .21. The results in this article should be regarded as preliminary until replicated. Nevertheless, the present results may reflect the fact that both PF and obsessing are cognitive mediated process, and the formal similarity led to higher correlations. This interpretation is especially likely for a nonselected student sample, where compulsions are less frequent than obsessions or obsession-like thoughts. In light of a recent research, the relationship between PF, obsessions, and compulsions is in need of further elucidation. On the one hand, our hypothesis was not confirmed, and PF did not explain OC symptoms above and beyond constructs of obsessive beliefs in a sample of students with high levels of OC symptomatology (Abramowitz, 2009). In contrast, ACT for obsessive-compulsive disorder has been found to distinctively decrease compulsions by promoting acceptance of obsessions (Twohig, Hayes, & Masuda, 2006; Twohig et al., 2010). This change was accompanied by an increase in PF. In order to ascertain whether PF is a salient mediator of treatment outcome, tightly controlled studies using other methods that allow for examination of temporal changes in obsessions, compulsions, obsessive beliefs, and PF are needed (Gloster et al., 2008; Twohig et al., 2006).

Perhaps the most important criticism of PF is that it is nothing substantially new (Hofmann & Asmundson, 2008). The recent empirical investigations into aspects of PF that failed to find indications of incremental validity (Abramowitz, 2009; Kollman et al., 2009) are all the more critical. To be sure, psychology is replete with constructs that have difficulty establishing incremental validity above and beyond that which came before (Haynes & Lench, 2003; Smith, Fischer, & Fister, 2003). In contrast, previous studies (Bond et al., in press) and the current results suggest a more qualified appraisal. Multiple regression analyses in our study showed that PF predicts variance above and beyond established constructs such as neuroticism, depressive symptomatology, and anxiety sensitivity. The predictive value varied, however, with respect to the outcome variable. Whereas PF incrementally predicted all outcomes of functioning and impairment in the PD/AG and SP samples (ΔR^2 = .012–.108), it did not do so with agoraphobic avoidance. This was surprising because it is a behavioral pattern conceptually incompatible to PF. At the symptom level, results were partially consistent with our hypotheses. On the one hand, PF did not significantly predict variance in panic symptoms above and beyond the predictors of anxiety sensitivity and depressive symptoms, but unique variance was predicted for HAM-A total scores (ΔR^2 = .022). No clear explanation accounts for these inconsistent findings. The comparatively expansive facets of the HAM-A (including some items that tap into functioning), in comparison with circumscribed panic symptoms, may have played a role. Recent results have also shown a close link between anxiety sensitivity and experiential avoidance in an analogue sample (Kelly & Forsyth, in press). If these results replicate, this would suggest that PF is most predictive of overarching functional level but that specific diagnostic measures are more useful in predicting symptoms. This can be viewed as partially consistent with theory in that PF emphasizes values and workability, both of which are likely more associated with functional level than symptom expression.

The results of this study must be interpreted with several limitations in mind. For example, some bordering constructs and parallel measures of interest were not included, and certain key levels of symptom severity could not be documented to allow for clean comparisons with previous studies (i.e., OCD symptomatology; Abramowitz, 2009). Specific measures designed to test construct validity were not included. Furthermore, PF was measured exclusively using the AAQ–II and, thereby, method variance was not systematically varied. That said, we are unaware of any established methods to measure PF other than self-report questionnaires, a format that may limit the sensitivity needed to capture

predictions made from PF. Future studies are encouraged to control for this using a multitrait–multimethod format. Finally, the diagnostic assessment of the SP sample was not independently verified by a clinician.

Despite these limitations, results from this study suggest that the AAQ–II can be regarded as a psychometrically sound measure of PF. This has been demonstrated for two larger clinically relevant help-seeking samples as well as for two nonclinical samples offering a broad basis for its empirically grounded application. The analyses of specific aspects of construct validity, namely clinical validity and incremental validity, reveal that PF is not only clinically useful but also adds unique, if overlapping, predictions and descriptions to those of more established constructs. This incremental contribution was importantly demonstrated with outcomes measuring functioning and impairment. The latter aspect seems important because these variables have been traditionally neglected, at least in comparison to the measurement, theorizing, and therapeutic targeting of symptoms. The present results indicate that PF can contribute to our understanding of psychological health, psychopathology, and therapy, in addition to addressing specific symptoms.

These generally positive findings are tempered by the fact that not all hypotheses were confirmed. The prediction that PF would correlate more strongly with compulsions than obsessions was exactly opposite, and the hypothesis that PF would explain additional variance in a differential pattern consistent with theory was not supported in all samples (i.e., the AAQ–II did not add explanatory value for all measures of avoidance and it did for some symptoms). Whereas these findings must be replicated in controlled studies before strong theoretical implications can be made, the results do document that predictions derived from PF theory are not so robust as to be immune to effects of sample and method variance. Furthermore, the reliable assessment of PF needs to extend beyond the self-report questionnaire into other formats, such as observation and implicit measurement. Nevertheless, there seems to be enough evidence among our various analyses that it is worthwhile to continue investigations into the construct PF. Future research is needed to address whether and under which conditions PF serves as a mediator of treatment outcome and to what degree it contributes to beneficial outcomes (Nelson-Gray, 2003), especially in terms of functioning and impairment and, preferably, across numerous therapeutic interventions and not just ACT studies.

References

Abramowitz, J. S., Lackey, G. R., & Wheaton, M. G. (2009). Obsessive–compulsive symptoms: The contribution of obsessional beliefs and experiential avoidance. *Journal of Anxiety Disorders, 23,* 160–166. doi:10.1016/j.janxdis.2008.06.003

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,* 176–181. doi:10.1037/1040-3590.10.2.176

Apfeldorf, W. J., Shear, M. K., Leon, A. C., & Portera, L. (1994). A brief screen for panic disorder. *Journal of Anxiety Disorders, 8,* 71–78. doi:10.1016/0887-6185(94)90024-8

Arch, J. J., & Craske, M. G. (2008). Acceptance and commitment therapy and cognitive behavioral therapy for anxiety disorders: Different treatments, similar mechanisms? *Clinical Psychology: Science and Practice, 15,* 263–279. doi:10.1111/j.1468-2850.2008.00137.x

Bach, P., & Hayes, S. C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 70,* 1129–1139. doi:10.1037/0022-006X.70.5.1129

Bandelow, B. (1995). Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The Panic and Agoraphobia Scale. *International Clinical Psychopharmacology, 10,* 73–82. doi:10.1097/00004850-199506000-00003

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory—II.* San Antonio, TX: Psychological Corporation.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Orcutt, H. K., Waltz, T., & Zettle, R. D. (in press). Preliminary psychometric properties of the Acceptance and Action Questionnaire—II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects*? Behaviour Research and Therapy, 41,* 1411–1426. doi:10.1016/S0005-7967(03)00059-7

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York, NY: Guilford Press.

Brown, T. A., Di Nardo, P. A., & Barlow, D. H. (1994). *Anxiety disorders interview schedule for DSM–IV (ADIS–IV).* San Antonio, TX: Psychological Corporation.

Chaker, S., Hofmann, S. G., & Hoyer, J. (2010). Can a one-weekend group therapy reduce fear of blushing? Results of an open trial. *Anxiety, Stress, & Coping: An International Journal, 23,* 303–318. doi:10.1080/10615800903075132

Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI) administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.

Ehlers, A., & Margraf, J. (1993). "Angst vor der Angst": Ein neues Konzept in der Diagnostik von Angststörungen [Fear of fear: A new concept in the assessment of anxiety disorders]. *Verhaltenstherapie, 3,* 14–24. doi:10.1159/000258732

Eifert, G. H., & Forsyth, J. P. (2005). *Acceptance & commitment therapy for anxiety disorders: A practitioner's treatment guide to using mindfulness, acceptance, and values-based behavior change strategies.* Oakland, CA: New Harbinger.

EuroQol Group. (1990). EuroQuol—A new faclility for the measurement of health-related quality of life. *Health Policy, 16,* 199–208. doi:10.1016/0168-8510(90)90421-9

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43,* 522–527. doi:10.1037/h0076760

Fernandéz-Ballesteros, R. (2004). Self-report questionnaires. In M. Herson (Series Ed.), S. N. Haynes, & E. M. Heiby (Vol. Eds.), *Comprehensive handbook of behavioral assessment: Vol. 3. Behavioral assessment* (pp. 194–221). New York, NY: Wiley.

Filipp, S.-H., & Freudenberg, E. (1989). *Der Fragebogen zur Erfassung dispositionaler Selbstaufmerksamkeit (SAM-Fragebogen)* [A questionnaire for the assessment of dispositional self-awareness (SAM questionnaire)]. Göttingen, Germany: Hogrefe.

Fletcher, L., & Hayes, S. C. (2005). Relational frame theory, acceptance and commitment therapy, and a functional analytic definition of mindfulness. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 23,* 315–336. doi:10.1007/s10942-005-0017-7

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299. doi:10.1037/1040-3590.7.3.286

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: De-

velopment and validation of a short version. *Psychological Assessment, 14,* 485–495. doi:10.1037/1040-3590.14.4.485

Foa, E. B., Kozak, M. J., Salkovskis, P. M., Coles, M. E., & Amir, N. (1998). The validation of a new obsessive-compulsive disorder scale: The obsessive-compulsive inventory. *Psychological Assessment, 10,* 206–214. doi:10.1037/1040-3590.10.3.206

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31,* 772–799. doi:10.1177/0145445507302202

Franke, G. H. (2000). *Brief symptom inventory von L. R. Derogatis (Kurzform der SCL-90-R)—Deutsche version* [Brief symptom inventory of L. R. Derogatis (Short form of the SCL-90-R)—German version]. Göttingen, Germany: Beltz Test.

Gaudiano, B. A. (2009). Öst's (2008) methodological comparison of clinical trials of acceptance and commitment therapy versus cognitive behavior therapy: Matching apples with oranges? *Behaviour Research and Therapy, 47,* 1066–1070. doi:10.1016/j.brat.2009.07.020

Geisheim, C., Hahlweg, K., Fiegenbaum, W., Frank, M., Schroder, B., & von Witzleben, I. (2002). The German version of the Brief Symptom Inventory (BSI): Reliability and validity in a sample of outpatient psychotherapy patients. *Diagnostica, 48,* 28–36. doi:10.1026//0012-1924.48.1.28

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11,* 572–580. doi:10.1086/208993

Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Acceptance-based treatment for smoking cessation. *Behavior Therapy, 35,* 689–705. doi:10.1016/S0005-7894(04)80015-7

Gloster, A. T., Richard, D. C. S., Himle, J., Koch, E., Anson, H., Lokers, L., & Thornton, J. (2008). Accuracy of retrospective memory and covariation estimation in patients with obsessive-compulsive disorder. *Behaviour Research and Therapy, 46,* 642–655. doi:10.1016/j.brat.2008.02.010

Gloster, A. T., Wittchen, H.-U., Einsle, F., Höfler, M., Lang, T., Helbig-Lang, S., & Arolt, V. (2009). Mechanism of action in CBT (MAC): Methods of a multi-center randomized controlled trial in 369 patients with panic disorder and agoraphobia. *European Archives of Psychiatry and Clinical Neuroscience, 259,* S155–S166. doi:10.1007/s00406-009-0065-6

Gloster, A. T., Wittchen, H.-U., Einsle, F., Lang, T., Helbig-Lang, S., Fydrich, T., . . . Arolt, V. (2011). Psychological treatment for panic disorder with agoraphobia: A randomized controlled trial to examine the role of therapist-guided exposure in-situ in CBT. *Journal of Consulting and Clinical Psychology, 79,* 406–420. doi:10.1037/a0023584

Gönner, S., Leonhart, R., & Ecker, W. (2008). The Obsessive-Compulsive Inventory—Revised (OCI–R): Validation of the German version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders, 22,* 734–749. doi:10.1016/j.janxdis.2007.07.007

Guy, W. (1976). *Clinical global impression.* In W. Guy (Ed.), *ECDEU assessment manual for psychopharmacology revised* (pp. 217–222.). National Institute of Mental Health: Rockville, MD.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology, 32,* 50–55.

Hammelstein, P., & Soifer, S. (2006). Is "shy bladder syndrome" (paruresis) correctly classified as social phobia? *Journal of Anxiety Disorders, 20,* 296–311. doi:10.1016/j.janxdis.2005.02.008

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI–II: Beck Depressionsinventar Manual* [Beck Depression Inventory Manual]. Frankfurt, Germany: Harcourt.

Hayes, S. C., Luoma, J., Bond, F., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes, and outcomes. *Behaviour Research and Therapy, 44,* 1–25. doi:10.1016/j.brat.2005.06.006

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change.* New York, NY: Guilford Press.

Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., . . . McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record, 54,* 553–578.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456–466. doi:10.1037/1040-3590.15.4.456

Herzberg, P. Y., & Brähler, E. (2006). Assessing the Big-Five personality domains via short forms: A cautionary note and a proposal. *European Journal of Psychological Assessment, 22,* 139–148. doi:10.1027/1015-5759.22.3.139

Hofmann, S. G., & Asmundson, G. J. G. (2008). Acceptance and mindfulness-based therapy: New wave or old hat? *Clinical Psychology Review, 28,* 1–16. doi:10.1016/j.cpr.2007.09.003

Hoyer, J., Ruhl, U., Scholz, D., & Wittchen, H.-U. (2006). Patients' feedback after computer-assisted diagnostic interviews for mental disorders. *Psychotherapy Research, 16,* 357–363. doi:10.1080/10503300500485540

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453. doi:10.1037/1082-989X.3.4.424

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27,* 197–209. doi:10.1901/jaba.1994.27-197

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review, 30,* 865–878. doi:10.1016/j.cpr.2010.03.001

Kashdan, T. B., & Steger, M. F. (2006). Expanding the topography of social anxiety: An experience-sampling assessment of positive emotions, positive events, and emotion suppression. *Psychological Science, 17,* 120–128. doi:10.1111/j.1467-9280.2006.01674.x

Kelly, M. M., & Forsyth, J. P. (in press). Associations between emotional avoidance, anxiety sensitivity, and reactions to an observational fear challenge procedure. *Behaviour Research and Therapy.*

Kollman, D. M., Brown, T. A., & Barlow, D. H. (2009). The construct validity of acceptance: A multitrait–multimethod investigation. *Behavior Therapy, 40,* 205–218. doi:10.1016/j.beth.2008.06.002

Köppe, E. (2001). *Glückliche Eltern-liebe Kinder? Auswirkung von Partnerschaft und psychischer Symptomatik der Eltern auf das Verhalten der Kinder* [Happy parents—Nice children? Impact of partnership and parental symptomatology on children's behavior] (Unpublished doctoral dissertation). Technische Universitaet Braunschweig, Braunschweig, Germany.

Levin, M., & Hayes, S. C. (2009). Is acceptance and commitment therapy superior to established treatment comparisons? *Psychotherapy and Psychosomatics, 78,* 380. doi:10.1159/000235978

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. doi:10.1016/0005-7967(94)00075-U

McCracken, L. M., Vowles, K. E., & Eccleston, C. (2005). Acceptance-based treatment for persons with complex, long standing chronic pain: A preliminary analysis of treatment outcome in comparison to a waiting phase. *Behaviour Research and Therapy, 43,* 1335–1346. doi:10.1016/j.brat.2004.10.003

Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coeffients. *Psychological Bulletin, 111,* 172–175. doi:10.1037/0033-2909.111.1.172

Muthén, B., & Muthén, L. (2008). *Mplus version 5 software users guide.* Los Angeles, CA: Author.

Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment, 15,* 521–531. doi:10.1037/1040-3590.15.4.521

Öst, L. G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy, 46,* 296–321. doi:10.1016/j.brat.2007.12.005

Peterson, R. A., & Reiss, S. (1993). *Anxiety sensitivity index revised test manual.* Worthington, OH: IDS Publishing Corporation.

Powers, M. B., Vörding, M. B., & Emmelkamp, P. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherapy and Psychosomatics, 78,* 73–80. doi:10.1159/000190790

Ravens-Sieberer, U., Wille, N., Badia, X., Bonsel, G., Burström, K., Cavrini, G., & Greiner, W. (2010). Feasibility, reliability, and validity of the EQ-5D-Y: Results from a multinational study. *Quality of Life Research, 19,* 887–897. doi:10.1007/s11136-010-9649-x

Raykov, T. (2001). Estimation of Congeneric Scale Reliability via covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54,* 315–323. doi:10.1348/000711001159582

Reznick, J. S., Hegeman, I. M., Kaufman, E. R., Woods, S. W., & Jacobs, M. (1992). Retrospective and concurrent self-report of behavioral inhibition and their relation to adult mental health. *Development and Psychopathology, 4,* 301–321. doi:10.1017/S095457940000016X

Rohrbacher, H., Hoyer, J., Beesdo, K., Hoefler, M., Bittner, A., Lieb, R., & Wittchen, H.-U. (2008). The Behavioral Inhibition Scale and its psychometric properties in a representative German sample. *International Journal of Methods in Psychiatric Research, 17,* 80–88. doi:10.1002/mpr.233

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Schmaltz, J. E., & Murrell, A. R. (2010). Measuring experiential avoidance in adults: The avoidance and fusion questionnaire. *International Journal of Behavioral Consultation and Therapy, 6,* 198–213.

Schneider, S., & Margraf, J. (2006). *Diagnostisches Interview bei psychischen Störungen* (*DIPS für* DSM–IV–TR) (Diagnostic interview for mental disorders [DIPS for *DSM–IV–TR*; 4th ed., text rev.]). Berlin, Germany: Springer.

Shear, M. K., Vander Bilt, J., Rucci, P., Endicott, J., Lydiard, B., Otto, M. W., . . . Frank, D. M. (2001). Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). *Depression and Anxiety, 13,* 166–178. doi:10.1002/da.1033

Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment, 15,* 467–477. doi:10.1037/1040-3590.15.4.467

Twohig, M. P., Hayes, S. C., & Masuda, A. (2006). Increasing willingness to experience obsessions: Acceptance and commitment therapy as a treatment for obsessive-compulsive disorder. *Behavior Therapy, 37,* 3–13. doi:10.1016/j.beth.2005.02.001

Twohig, M. P., Hayes, S. C., Plumb, J. C., Pruitt, L. D., Collins, A. B., Hazlett-Stevens, H., & Woidneck, M. R. (2010). A randomized clinical trial of acceptance and commitment therapy versus progressive relaxation training for obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology, 78,* 705–716. doi:10.1037/a0020508

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg [A revised version of the German Rosenberg Self-Esteem Scale]. *Zeitschrift für Differentielle und DiagnostischePsychologie, 24,* 3–7. doi:10.1024//0170-1789.24.1.3

von der Schulenburg, J.-M., Claes, C., Greiner, W., & Uber, A. (1998). The German version of the EuroQol Questionnaire. *Zeitschrift für Gesundheitswissenschaft, 6,* 3–20.

Wittchen, H.-U., & Pfister, H. (1997). *DIA-X interviews: Manual für Screeningverfahren und Interview; Interviewheft Längsschnittuntersuchung (DIA-X lifetime); Interviewheft Querschnittsuntersuchung (DIA-X-12 Monate); PC Programm zur Durchführung des Interviews; Auswertungsprogramm* [Manual for screening procedures and interview; interview booklet for longitudinal assessments (DIA-X Lifetime); interview booklet for cross sectional assessments (DIA-X 12 months); PC-program for the administration of the interview and analyses]. Frankfurt am Main, Germany: Sweets & Zeitlinger.

World Health Organization. (1990). *Composite international diagnostic interview (CIDI), Version 1.0.* Geneva, Switzerland: World Health Organization.

Zaider, T. I., Heimberg, R. G., Fresco, D. M., Schneier, R. R., & Liebowitz, M. R. (2003). Evaluation of the Clinical Global Impression Scale among individuals with social anxiety disorder. *Psychological Medicine, 33,* 611–622. doi:10.1017/S0033291703007414

Zettle, R. D., & Hayes, S. C. (1986). Dysfunctional control by client verbal behavior: The context of reason giving. *The Analysis of Verbal Behavior, 4,* 30–38.

Zettle, R. D., & Rains, J. C. (1989). Group cognitive and contextual therapies in treatment of depression. *Journal of Clinical Psychology, 45,* 436–445. doi:10.1002/1097-4679(198905)45:3<436::AID-JCLP2270450314>3.0.CO;2-L

Received July 9, 2010
Revision received April 15, 2011
Accepted April 20, 2011 ■