Journal of Psychopathology and Behavioral Assessment (2025) 47:24
https://doi.org/10.1007/s10862-025-10200-4

A Comparison of the Long and Short Forms of the Penn State Worry Questionnaire in Adults with Generalised Anxiety Disorder

Emily J. Wilson[1,2] · Maree J. Abbott[2,6] · Alice R. Norton[2] · David Berle[3] · Lexine Stapinski[4] · Ronald M. Rapee[5]

Accepted: 2 February 2025 / Published online: 22 February 2025

Abstract

The Penn State Worry Questionnaire (PSWQ) is a 16-item self-report measure designed to assess pathological worry. The PSWQ has, however, demonstrated inconsistent factor structure in adults and older adults leading to the development of the 8-item PSWQ-A and the ultra-brief 3-item PSWQ-3. The PSWQ is yet to be compared to the PSWQ-A and PSWQ-3 in adults with generalised anxiety disorder (GAD). Thus, the current study aimed to evaluate the psychometric properties of these three versions. Participants were screened using the Anxiety Disorders Interview Schedule for DSM-IV-TR to ascertain clinical principal diagnosis of GAD ($n=140$) or non-clinical status ($n=76$). Four different confirmatory factor models were fit to the 16-item PSWQ, with a unidimensional model fit to the 8-item PSWQ-A and to the PSWQ-3. A bifactor model fitted the data best for the PSWQ, and a unidimensional PSWQ-A model fitted the data best for the GAD sample. Results found that all three versions of the PSWQ demonstrated good construct validity, moderate test-retest reliability, and excellent criterion validity. ROC curve analysis indicated that all three versions demonstrated comparable levels of sensitivity and specificity for screening GAD. Both the PSWQ-A and PSWQ demonstrated no floor or ceiling effects and good internal consistency, whereas the PSWQ-3 demonstrated floor effects with adequate internal consistency. Overall, all three versions of the PSWQ share comparable psychometric properties. As such, the brevity of the PSWQ-A and its comparable performance to the 16-item PSWQ, warrant recommendations for use of this version to researchers and clinicians.

Keywords Generalised Anxiety Disorder · Worry · Confirmatory Factor Analysis · Psychometrics · Self-report

Introduction

Worry is a form of repetitive negative thinking (McEvoy et al., 2019a) that is conceptualised as a mental process involving attempts to plan and prepare a favourable solution in the face of an uncertain and potentially negative outcomes (Borkovec, 1994; Fresco et al., 2002). Though worry is experienced to some extent by everyone, excessive, multifocal, and difficult to control worry is a cardinal feature of generalised anxiety disorder (GAD). In GAD, worry is typified by statements that imply catastrophising or inflexible rule-bound interpretations (Dugas et al., 1998; Molina et al., 1998). Worry content centres around life domains such as family, interpersonal relationships, finances, personal health, health of loved ones, work, education, everyday tasks, the world and society, and is relatively stable over time (Constans et al., 2002). In particular, worry themes tend to be related to areas of high personal value (Boehnke et al., 1998).

✉ Maree J. Abbott
maree.abbott@sydney.edu.au

1 Graduate School of Health, University of Technology Sydney, Ultimo, Australia

2 School of Psychology, The University of Sydney, Sydney, Australia

3 School of Medicine and Psychology, Australian National University, Canberra, Australia

4 Sydney Medical School, The Matilda Centre, The University of Sydney, Sydney, Australia

5 School of Psychological Sciences, Macquarie University, Sydney, Australia

6 Brain and Mind Centre, School of Psychology, The University of Sydney, 94 Mallett Street Camperdown, Sydney, New South Wales 2050, Australia

The Penn State Worry Questionnaire is a 16-item self-report questionnaire that is one of the most widely used measures of excessive worry (McEvoy et al., 2019b). The PSWQ has demonstrated good construct validity, internal consistency, and test-retest reliability (Brown et al., 1992; Dear et al., 2011; Hazlett-Stevens et al., 2004; Meyer et al., 1990). Despite these favourable psychometric properties, there has been inconsistent findings in relation to factor structure. The PSWQ was originally found to have a unidimensional factor structure in both undergraduates (Meyer et al., 1990) and in a heterogenous anxiety disorder sample (Brown et al., 1992). In contrast, Fresco et al. (2002) found that a two-factor solution provided superior fit when compared to a unidimensional solution in undergraduate students, with these two factors comprising of (1) worry engagement and (2) absence of worry. This same study found that higher order and lower order factors, general worry and worry engagement respectively, explained the majority of variance in symptom measures (Fresco et al., 2002). Thus, the second factor, absence of worry, appeared to represent methodological variance, as the items loaded onto this factor were negatively worded (i.e., *'I do not tend to worry about things')* and thus reverse scored, rather than representing a conceptually orthogonal construct. Further investigation in a mixed sample of people with anxiety and mood disorders found that a unidimensional model accounting for method effects (i.e., covariance among errors of reverse-scored items) provided significantly better fit compared to the two-factorsolution (Brown, 2003). Therefore, it appears that the negatively phrased items may have an aberrant function producing an artificial factor, creating a limitation of the 16-item PSWQ. These findings also parallel literature that discourages the intermixing of negatively worded items in scales as they create psychometric artefacts, rather than reducing respondent bias (Chyung et al., 2018; Roszkowski & Soven, 2010), with evidence indicating that reverse-scored items hinder psychometric performance in other anxiety disorder questionnaires (Rodebaugh et al., 2007).

In light of the inconsistent factor structure of the full PSWQ, the 8-item PSWQ-Abbreviated (PSWQ-A) was constructed by Hopko et al. (2003). The PSWQ-A was developed from the full PSWQ items that were only positively worded (i.e., *'I worry all the time'*) to address factorial redundancy as well difficulties that had been previously reported for older adults with answering negatively worded questions due to increased cognitive load. Hopko et al. (2003) demonstrated that the PSWQ-A has strong fit indices, good convergent validity, high internal consistency, and adequate test–retest reliability in a sample of older adults with a principal or co-principal diagnosis of GAD. Wuthrich et al. (2014) went on to reproduce this methodology and compare the PSWQ to the PSWQ-A with a clinical sample of older adults with depression and an anxiety disorder. Wuthrich et al. (2014) again found poor fit across absolute and incremental fit indices for both the one-factor and two-factor models for the 16-item PSWQ, whereas the unidimensional PSWQ-A was found to have good fit across all indices. Further, the PSWQ-A was found to have good construct validity and internal consistency, with adequate test-retest reliability. In addition to the PSWQ-A, another 3-item version of the PSWQ has also been developed by Berle and colleagues (2011). The PSWQ-3 was created to incorporate the essential features of pathological worry (i.e., high frequency, high uncontrollability, multiple worry domains) as well as to exclude any reverse-scored items (Berle et al., 2011). Psychometric properties of the PSWQ-3 were explored with a mixed sample of adults with a principal anxiety or related disorder, and the results indicated similar psychometric properties for the 16-item PSWQ and PSWQ-3. Berle et al. (2011) found comparable construct validity, as well as sensitivity to treatment for individuals with GAD with equivalent large effect sizes following individual cognitive-behaviour therapy (CBT). Participants with GAD also scored higher than participants with another principal anxiety disorder on both the PSWQ and PSWQ-3.

Kertz et al. (2014) aimed to compare all three versions of the PSWQ in a heterogeneous clinical sample (i.e., mood, anxiety, trauma-related, and psychotic disorders) of adults presenting for treatment in a partial hospital setting. The underlying factor structure of the 16-item PSWQ was represented by a two-factor model, although, when compared to the PSWQ-A, the PSWQ-A exhibited the best fit across all indices (i.e., SB χ^2, RMSEA, CFI, GFI and SRMR). In addition, all three versions demonstrated similar psychometric properties, with good construct validity for the briefer versions and excellent internal consistency for the full PSWQ. All three versions showed comparable sensitivity to treatment, demonstrating medium effects following an assorted program of group and individual therapy informed by cognitive behaviour therapy. In adults with GAD, only Dear et al. (2011) has compared four model iterations of the 16-item PSWQ (i.e., unidimensional, two-factor, one-factor with method effects, and a three-factor model) and investigated the psychometric properties in adults with GAD. Interestingly, Dear et al. (2011) found a three-factor solution, with all items loading onto one general factor as well as two separate method factors (i.e., absence of worry and worry engagement), provided the best fit to the data. The path diagram of the three-factor model from Dear et al. (2011, p.20), demonstrated a similar structure to a bifactor model. Bifactor modelling has an advantage (over the three-factor solution) as bifactor indices can determine the extent to which a unidimensional (or multidimensional)

interpretation is supported by the data. In addition, research has yet to compare the 16-item PSWQ to the two shortened versions, PSWQ-A and PSWQ-3, in a clinical sample of *adults* with a *principal diagnosis of GAD*.

The current study aimed to extend previous research by comparing the psychometric properties of the 16-item PSWQ, 8-item PSWQ-A, and 3-item PSWQ-3 in a sample of adults with a principal diagnosis of GAD. It was hypothesised that the shorter PSWQ-A unidimensional model would provide the best fit for the data, when compared to the 16-item PSWQ and 3-item PSWQ-3 (Kertz et al., 2014). In addition, we hypothesised that both PSWQ-A and PSWQ-3 would demonstrate comparable psychometric properties to the longer PSWQ-16. Specifically, we predicted that all three versions would demonstrate no floor or ceiling effects. All three versions were predicted to demonstrate good construct validity though significant positive moderate correlations with a GAD symptom measures (i.e., physiological tension) and processes hypothesised to maintain pathological worry in cognitive-behavioural models of GAD (i.e., intolerance of uncertainty and negative metacognitive beliefs about worry). Significant moderate positive correlations were predicted for these aforementioned variables in line with previous research in mixed clinical and undergraduate samples (Kertz et al., 2014; Wells & Cartwright-Hatton, 2004). We also predicted significant low correlations with distinct, yet overlapping psychopathology constructs, including depression and autonomic anxiety, as well as a range of metacognitive beliefs (i.e., positive beliefs about worry, lack of cognitive confidence, cognitive self-consciousness, and need for control of thoughts) in line with previous research (Kertz et al., 2014; Wells & Cartwright-Hatton, 2004). Further, all three versions were predicted to demonstrate good internal consistency as well as reproducibility through adequate test-retest reliability over a 12-week period. Finally, we also predicted that the measure would significantly discriminate participants with GAD from non-clinical participants, with high sensitivity and specificity.

Method

Participants

A total of 216 participants were included in the study, with the combined group being predominantly female (73.6%), with an age range from 18 to 70 years (M=36.58 years, SD=13.07). All participants were assessed using the Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV; Brown et al., 1994). Participants needed to be at least 18 years old, living in Australia and fluent in English to participate. Individuals who were experiencing active psychosis or active suicidal ideation were excluded from the study and referred to appropriate mental health services.

Non-clinical participants (N=76) were predominantly female (65.8%) and aged between 18 and 70 years (M=35.55; SD=14.08). Non-clinical participants were screened using the using the ADIS-IV and did not meet diagnostic criteria for any mental health disorder. Non-clinical participants were recruited via community notices in local newspapers.

The clinical sample was recruited from a specialist university-based clinical research unit for the assessment and treatment of anxiety disorders. All clinical participants met criteria for a principal diagnosis of GAD (N=140) and were aged between 18 and 70 years old (M=37.14; SD=12.50), with the majority identifying as female (77.9%). The clinician severity rating (CSR) was assigned by the clinician who administered the ADIS-IV and reflects the severity of symptoms and related interference or distress. The CSR ranges from zero to eight, a CSR of four or higher suggests a clinical level of severity. The average CSR for the principal diagnosis of GAD in this study was 6.15 (SD=0.93). Regarding co-morbidity within the clinical sample, 73.6% of participants satisfied criteria for at least one additional diagnosis (i.e., secondary diagnosis), including social anxiety disorder (41.4%), major depressive disorder (12.9%), dysthymia (5.0%), panic disorder with agoraphobia (5%), panic disorder without agoraphobia (1.4%), obsessive-compulsive disorder (1.4%), specific phobia (2.9%), and other (3.6%). At the time of assessment, 41.4% reported that they were taking some form of psychotropic medication.

There were no significant differences between the two groups on age, t (214)=0.82, p=.41, gender χ^2 (1)=3.69, p=.06, employment status, χ^2 (7)=4.76, p=.68 and education, χ^2 (2)=57, p=.06. There was a significant difference between groups in terms of marital status, χ^2 (4)=24.56, p<.001. A greater proportion of the non-clinical group were single (52.6%) or had been divorced (18.4%), compared to the clinical group who were single (38.7%) or divorced (6.4%). Whereas in the clinical group the majority were married/de facto (54.9%) compared to the non-clinical group (27.6%). With only a small percentage of people widowed in the non-clinical sample (1.4%).

Procedure

The original research study was approved by the Macquarie University Human Research Ethics Committee (Project HE-R02594), and the methodology of this study was approved by The University of Technology Sydney Human Research Ethics Committee (Project: ETH22-7702). All participants (N=140) completed the measures as part of a

clinical trial at the initial assessment session (Abbott, 2007). At the initial assessment session, all participants signed the consent form and were administered the ADIS-IV by a clinical psychologist or doctoral-level graduate student, supervised by senior clinical psychologists. A subsample ($N=22$) of the participants were in a waitlist condition for the clinical trial. These participants re-completed the PSWQ after a 12-week waitlist period (with no treatment). Participants in the waitlist condition were offered either group cognitive behaviour therapy or group mindfulness-based therapy after the 12-week waitlist, with results from the treatment not reported in the current study.

Measures

Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV)

The ADIS-IV is a semi-structured interview designed to evaluate anxiety and mood disorders in accordance with DSM-IV diagnostic criteria (Brown et al., 1994). Extensive research has demonstrated good to excellent inter-rater reliability for this interview (Brown et al., 1994). In the current study, blind raters exhibited high agreement ($\kappa=0.84$), underscoring strong inter-rater reliability in assessing diagnostic consistency.

Penn State Worry Questionnaire (PSWQ)

The PSWQ is a measure of excessive worry (Meyer et al., 1990). The PSWQ is a 16-item inventory created to capture the excessive and uncontrollable nature of pathological worry. The PSWQ is rated on a 5-point Likert scale, with scores ranging from 16 to 80. There are 12 items that are positively worded (i.e., '*I worry all the* time') and the remaining five items are negatively worded (i.e., '*I never worry about* anything') and need to be reverse scored. The PSWQ has been found to have good construct and internal reliability in a sample of participants with GAD (Brown et al., 1992; Dear et al., 2011). Internal consistency for the combined sample was $\alpha=0.97$.

Penn State Worry Questionnaire- Abbreviated (PSWQ-A)

The PSWQ-A is an abridged version of the PSWQ, incorporating only eight positively worded items from the original PSWQ (Hopko et al., 2003). The PSWQ-A has been found to have a stable one-factor structure in older adult samples with mixed anxiety as well as undergraduate samples (Crittendon & Hopko, 2006; Wuthrich et al., 2014). The PSWQ-A was derived from the PSWQ and was not provided as a separate measure to participants. Cronbach's α for the combined sample was 0.95.

Penn State Worry Questionnaire – 3 (PSWQ-3)

The PSWQ-3 is an ultra-brief version on the PSWQ, with three items selected from the original PSWQ (Berle et al., 2011). The three items reflect the cardinal features of pathological worry (i.e., high frequency, perceived uncontrollability and multiple domains of worry). The PSWQ-3 has been found have good psychometric properties in both adults with GAD, as well as heterogenous mental health samples (Berle et al., 2011; Kertz et al., 2014). Like the PSWQ-A, the PSWQ-3 was extracted from the PSWQ and was not provided as a separate measure to participants. Cronbach's α for the combined sample was 0.93.

Intolerance of Uncertainty Scale – 12 (IUS – 12)

The IUS-12 was used as a measure of intolerance of uncertainty (Carleton et al., 2007; Freeston et al., 1994). There are 12 items that assess specific negative beliefs about uncertainty that someone may hold, such as '*When it is time to act*, *uncertainty paralyses me*' or '*I always want to know what the future has in store for me'*. The IUS-12 has a bifactor structure, with the total score to be utilised in the clinical sample of individuals with GAD (Wilson et al., 2020). It has also demonstrated good construct validity, treatment sensitivity, test-retest reliability as well as good internal consistency in a sample of individuals with GAD (Wilson et al., 2020). Internal consistency for the combined sample was excellent ($\alpha=0.93$).

Metacognitions Questionnaire-30 (MCQ-30)

The MCQ-30 is 30-item inventory, developed to assess metacognitive beliefs and cognitive monitoring tendencies (Wells & Cartwright-Hatton, 2004). There are five subscales in the MCQ-30: (1) cognitive self-consciousness; (2) beliefs about the need to control thoughts; (3) positive beliefs about worry; (4) cognitive confidence; and (5) negative beliefs about the uncontrollability and danger of worry. The MCQ-30 has demonstrated acceptable fit within a five-factor model, as well as good construct validity and internal consistency for each subscale (Huntley et al., 2020; Nordahl et al., 2023). In the current sample, internal consistency ranged from adequate to excellent in the combined sample: cognitive confidence, $\alpha=0.90$; positive beliefs about worry, $\alpha=0.90$; cognitive self-consciousness, $\alpha=0.85$; uncontrollability of thoughts and danger, $\alpha=0.91$; and beliefs about need to control thoughts, $\alpha=0.77$.

Depression Anxiety Stress Scales – Short Form (DASS-21)

The DASS-21 (Lovibond & Lovibond, 1995) includes 21-items aimed at assessing current symptoms of depression, anxiety (e.g., autonomic arousal and situational anxiety), and stress (e.g., agitation, impatience and nervous tension). Factor analysis has consistently demonstrated a three-factor solution, as well as adequate construct validity, internal consistency, and test-retest reliability in a variety of clinical samples (Brown et al., 1997; Lovibond & Lovibond, 1995; Page et al., 2007). Cronbach's α for each of the subscales was good to excellent for the combined sample: depression α = 0.91; anxiety α = 0.82; and stress α = 0.92.

Data Analysis

Confirmatory factor analysis (CFA) was performed to ascertain the best model that fit the data for the 16-item PSWQ, and 8-item PSWQ-A. CFA was selected (over EFA) as there were strong a priori hypotheses regarding the different factor structures from previous research (Flora & Flake, 2017). Models tested for the 16-item PSWQ were informed by previous research (Brown et al., 1997; Dear et al., 2011) and included: (1) a one-factor model with all items loaded onto one general worry factor; (2) a one-factor model accounting for method effects, with all items loaded onto one general worry factor, with an additional method factor to account for the residual covariations among responses to five negatively worded items (i.e., reverse-scored items were permitted to covary with one another, items 1, 3, 8, 10 and 11); (3) a two-factor model with the positively-worded items loaded onto one factor, and the five negatively worded items loaded onto a second factor; (4) bifactor model (i.e., similar to the three factor model tested by Dear et al., 2011) with one general factor (shared variance between items) and two group factors, positive-scored factor and negatively scored (indicative of the covariance not explained by the general factor). The correlations between the general worry factor and the two factors (i.e., positively scored and reverse-scored factors) were constrained to zero. A unidimensional model was fit to the 8-item PSWQ-A and to the 3-item PSWQ-3.

Each confirmatory factor model was created using polychoric correlation coefficients in Mplus, version 6.1 (Muthen & Muthen, 1998–2022). For each CFA model absolute and incremental goodness-of-fit indices were used to evaluate fit (Hu & Bentler, 1999). Missing data was managed using the full information maximum likelihood method (FIML) with a bootstrap resample of 1000 (Hayes, 2009). Absolute goodness-of-fit was evaluated using the chi-squared statistic, where a non-significant value indicates acceptable fit (Tabachnick & Fidell, 2013). Further absolute goodness-of-fit was indicated with the standardised root mean square residual (SRMR), the root mean square error of approximation (RMSEA), as well as the accompanying RMSEA 90% confidence interval (RMSEA 90% CI) and p of close fit (PCLOSE). For both, SRMR and RMSEA values ≤ 0.08 are indicative of acceptable fit, and values < 0.06 have been suggested as indicative of good fit (Hu & Bentler, 1999). A non-significant PCLOSE and lower limit of the RMSEA 90% CI close to zero indicate good model fit (Kenny et al., 2014). Incremental fit was evaluated with the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), and comparative fit index (CFI; Bentler, 1990). Values ≥ 0.90 for the TLI and CFI are suggestive of good fit, with values for the TLI and CFI ≥ 0.95 suggestive of excellent model fit (Hu & Bentler, 1999). Standardised factor loadings were also reported.

The bifactor indices were computed using the Bifactor Indices Calculator (Dueber, 2017) to determine the extent to which a unidimensional interpretation of PSWQ is supported by the data. Coefficient omega (ω) and omega subscale (ω_S) is a model-based estimate of internal reliability that can be applied to the general factor and group factors, and represents the proportion of variance in raw scores for the total score (ω) and each subscale (ω_S) that is explained by all sources of common variance i.e., both general and each subscale factor (Rodriguez et al., 2016). OmegaH (ω_H) represents the proportion of variance in the total score explained by the general factor. When OmegaH ≥ 0.80 the total score reflects a single construct sufficiently well (Rodriguez et al., 2016). OmegaHS (ω_{HS}), is an index reflecting the reliability of a subscale score (or the unique variance of each group factor) after controlling for the variance accounted for by the general factor (Reise et al., 2013). H is a measure of construct replicability conceptualized by and represents the correlation between a factor and an optimally-weighted item composite (Hancock & Mueller, 2001), thus high H values (> 0.80) suggest a well-defined latent variable. Specifically, explained common variance (ECV) reflects the proportion of all common variance explained by the general factor relative to the group factors. ECV > 0.70 indicates that a measure is sufficiently unidimensional to be treated as such in a latent variable modeling framework (Rodriguez et al., 2016). The Percent of Uncontaminated Correlations (PUC) represents the percentage of covariance terms which only reflect variance from the general dimension, thus the higher the PUC, the more the correlation matrix reflects the general factor (Rodriguez et al., 2016).

Subsequent psychometric analyses were conducted in SPSS 28.0.0 across all three measures. Data was screened for missing values Little's Missing Completely at Random Test and indicated that the data was missing completely at random, $\chi^2(45) = 46.03$, $p = .43$. Missing data was handled using a multiple imputation method. The Shapiro-Wilk statistic was non-significant ($p > .05$) for the 16-item PSWQ,

the 8-item PSWQ-A and PSWQ-3, indicating normality of data. Inspection of data and histograms indicated acceptable levels of skewness (i.e., < 2) and kurtosis (i.e., < 7). Group mean differences were calculated between the GAD and non-clinical samples using a one-way ANOVA. Floor and ceiling effects were considered to be present if 15% or more of participants scored the lowest or highest score, respectively (Terwee et al., 2007). Construct validity was assessed for the clinical sample using Pearson's *r* correlations, with descriptors used as per the following ranges: low = 0.10 −0.39; moderate = 0.40-0.69; strong = 0.70 −0.90. (Akoglu, 2018). Internal consistency was examined using Cronbach's alpha, with a score between 0.70 and 0.95, deemed as adequate for quality criteria (Mokkink et al., 2010). Test-retest reliability was assessed using a subsample of the participants with GAD ($N = 22$) who were in a waitlist condition for the clinical trial. Specifically, scores taken at the initial assessment were compared to scores taken at the end of the waitlist period (i.e., 12-weeks after the initial assessment). Test-retest reliability was assessed using the intraclass correlation coefficient (ICC). For the ICC (calculated for absolute agreement using the two-way random effects model) a value lower than 0.5 indicates poor reliability, a value between 0.5 and 0.75 suggests moderate reliability, a value between 0.75 and 0.90 indicates good reliability, with values greater than 0.9 suggesting excellent reliability (Koo & Li, 2016; McGraw & Wong, 1996). To assess criterion validity, the 16-item PSWQ will be correlated with the two briefer versions, PSWQ-A and PSWQ-3, using a split-half sample of participant with GAD. A Pearson's *r* correlation > 0.70 with the gold-standard measure (i.e., 16-item PSWQ) is suggested to meet quality criteria for criterion validity (Terwee et al., 2007). In addition, ROC curve analyses were conducted to determine if the measure could distinguish individuals with GAD from non-clinical participants. Specifically, for these analyses we reported the Youden-Index for the Area Under the Curve (AUC) which provides a cut-off score that maximises both sensitivity (proportion of true positive) and specificity (proportion of false negatives) (Fluss et al., 2005; Youngstrom, 2014). An accurate cut-off score should have an AUC close to 1.0, with scores above 0.85 classified as convincing evidence, scores between 0.75 and 0.85 classified as partially convincing, and AUC < 0.75 as unconvincing evidence for the accuracy of the cut-off score (Bowers & Zhou, 2019).

Results

Confirmatory Factor Analysis

Fit Indices

The results of the testing goodness of fit for the separate CFA models can be found in Table 1. See Table 2 for standardised factor loadings for the version of the PSWQ and PSWQ-A. The absolute and incremental fit indices were inconsistent for all four separate CFA models for the 16-item PSWQ. The bifactor model, though exhibited the most reliable fit for the 16-item PSWQ, however, there was room for improvement as the TLI < 0.90 and $\chi^2 p$ value was < 0.05. The PSWQ-A fit the data well across all absolute and incremental fit indices. A unidimensional model was also estimated for the 3-item PSWQ-3, however, the model was fully saturated, which was similar to (Kertz et al., 2014). As a result, goodness of fit indices could not be interpreted for the PSWQ-3.

Bifactor Indices

The omega coefficients for the general worry factor, and the positive worded factor were high. OmegaH for the general factor (ω_H) suggested that 80% of the variance in PSWQ scores can be explained by individual differences on the general factor. Whereas the omegaHS for the group factors (ω_{HS}) suggested only 13% and 29% of the variance in PSWQ scores can be attributed to the positively worded and negatively worded factors, respectively. Regarding the *H* construct, general factor had a high value (0.92), suggesting

Table 1 Fit indices for the PSWQ and PSWQ-A for adults with GAD

Fit Indices	PSWQ One Factor	PSWQ One Factor with method effects†	PSWQ Bifactor	PSWQ Two Factor	PSWQ-A One Factor
χ^2 / (*df*)	187.34 / 104	166.73 / 94	138.24 / 88	176.92 / 103	27.40 / 20
$\chi^2 p$ value	< 0.01	< 0.01	< 0.01	< 0.01	0.12
RMSEA [90% CI]	0.08 [0.06- 0.09]	0.07 [0.06- 0.09]	0.06 [0.04- 0.08]	0.07 [05- 0.09]	0.05 [< 0.01 −0.01]
PCLOSE	0.01	0.02	0.13	0.03	0.44
SRMR	0.07	0.06	0.05	0.06	0.05
CFI	0.86	0.88	0.92	0.88	0.97
TLI	0.84	0.84	0.88	0.86	0.95

Note. *PSWQ* Penn State Worry Questionnaire, *PSWQ-A* PSWQ-Abbreviated, *RMSEA* root mean square error of approximation, *PCLOSE* significance of the RMSEA close fit, *SRMR* standardised root mean square residual, *CFI* comparative fit index, *TLI* Tucker-Lewis index

† Method effects indicate that the reverse-scored items were permitted to covary with one another (items 1, 3, 8, 10 and 11)

Table 2 Standardised factor loadings for PSWQ, PSWQ-A, and PSWQ-3 as well as bifactor indices for the PSWQ for adults with GAD

Items	PSWQ One Factor	PSWQ One Factor with method effects†	PSWQ Two Factor		PSWQ Bifactor			PSWQ-A	PSWQ-3
			Positive	Negative	General	Positive	Negative		
1. If I do not have enough time to do everything, I do not worry about it.	0.40	0.39	-	0.47	0.38	-	0.22	-	-
2. My worries overwhelm me.	0.58	0.58	0.58	-	0.58	0.01	-	0.55	-
3. I do not tend to worry about things.	0.23	0.20	-	0.41	0.17	0.28	0.60	-	-
4. Many situations make me worry.	0.70	0.70	0.70	-	0.67	0.28	-	0.75	0.64
5. I know I should not worry about things; but I just cannot help it.	0.55	0.55	0.55	-	0.54	0.12	-	0.61	-
6. When I am under pressure I worry a lot.	0.55	0.55	0.55	-	0.53	0.21	-	0.60	-
7. I am always worrying about something.	0.81	0.81	0.81	-	0.76	0.30	-	0.79	-
8. I find it easy to dismiss worrisome thoughts.	0.30	0.29	-	0.32	0.32	-	0.10	-	-
9. As soon as I finish one task, I start to worry about everything else I have to do.	0.61	0.61	0.61	-	0.59	0.19	-	0.59	-
10. I never worry about anything.	0.50	0.47	-	0.66	0.45	-	0.50	-	-
11. When there is nothing more I can do about a concern, I do not worry about it anymore.	0.36	0.34	-	0.43	0.35	-	0.21	-	-
12. I have been a worrier all my life.	0.38	0.38	0.38	-	0.37	0.06	-	0.32	-
13. I notice that I have been worrying about things.	0.48	0.48	0.48	-	0.48	0.02	-	0.44	-
14. Once I start worrying; I cannot stop.	0.63	0.63	0.62	-	0.91	0.85	-	-	0.62
15. I worry all the time.	0.82	0.83	0.83	-	0.80	0.20	-	-	0.98
16. I worry about projects until they are all done.	0.61	0.62	0.62	-	0.58	0.25	-	-	-
Bifactor Indices									
Omega					ω = 0.90	ω_S = 0.91	ω_S = 0.60		
OmegaH					ω_H = 0.80	ω_{HS} = 0.13	ω_{HS} = 0.29		
H					0.92	0.79	0.50		
ECV					0.72				
PUC					0.37				

Note. *PSWQ* Penn State Worry Questionnaire, *PSWQ-A* PSWQ-Abbreviated, *ECV* Explained Common Variance, *PUC* Percent of Uncontaminated Correlations

† Method effects indicate that the reverse-scored items were permitted to covary with one another (items 1, 3, 8, 10 and 11)

a well-defined latent construct. Whereas, both positively worded and negatively worded factors were <0.80, suggesting that they are poorly defined latent constructs. The ECV indicates that the general factor explained 72% of the common variance, whereas only 27% was shared among the positively worded and negatively worded factors. The PUC indicated that 37% of the correlation matrix reflected the general factor.

Group Differences

Three separate independent samples *t*-test were performed on all three versions of the PSWQ, with each on finding that the clinical GAD group scored significantly higher those in the non-clinical group: PSWQ, $t\,(214) = 26.76$, $p < .001$; PSWQ-A, $t\,(214) = 24.51$, $p < .001$. PSWQ-3, $t\,(214) = 23.98$, $p < .001$. See Table 3 for descriptives.

Table 3 Group differences between non-clinical participants and adults with GAD for the three versions of the PSWQ

	PSWQ		PSWQ-A		PSWQ-3	
	Non-clinical (*N* = 78)	GAD (*N* = 140)	Non-clinical (*N* = 78)	GAD (*N* = 140)	Non-clinical (*N* = 78)	GAD (*N* = 140)
Mean (SD)	32.46 (9.70)	66.98 (8.68)	13.97 (5.75)	32.84 (5.20)	4.24 (1.62)	11.57 (2.38)

Note. *PSWQ* Penn State Worry Questionnaire, *PSWQ-A* PSWQ-Abbreviated

Table 4 Construct validity indicated by Pearsons' r correlations between the three versions of the PSWQ and associated GAD symptoms and processes in adults with GAD (*N* = 140)

Questionnaires	PSWQ	PSWQ-A	PSWQ-3
IUS – 12	0.43**	0.43**	0.40**
DASS – 21: Stress	0.57**	0.53**	0.55**
DASS – 21: Depression	0.26**	0.24**	0.28**
DASS – 21: Anxiety	0.32**	0.32**	0.34**
MCQ – 30: Negative beliefs about worry	0.55**	0.50**	0.59**
MCQ – 30: Positive beliefs about worry	0.29**	0.26**	0.18*
MCQ – 30: (Lack of) cognitive confidence	0.23**	0.21*	0.24**
MCQ – 30: Need for control of thoughts	0.32**	0.27**	0.31**
MCQ – 30: Cognitive self-consciousness	0.34**	0.28**	0.31**

Note. *PSWQ* Penn State Worry Questionnaire, *PSWQ-A* PSWQ-Abbreviated, *IUS-12* Intolerance of Uncertainty Scale −12, *DASS-21* Depression Anxiety Stress Scales-21, *MCQ-30* Metacognitive Beliefs Questionnaire-30

* $p < .05$. ** $p < .001$

Floor and Ceiling Effects

Across all 216 participants, scores on the 16-item PSWQ ranged from 16 to 80, with 0.5% achieving the lowest score of 16, and 1.4% achieved the highest possible score of 80. Across the combined sample, the 8-item PSWQ-A scores ranged from 8 to 40, with 3.2% achieving the lowest score of 8, and 2.8% achieved the highest possible score of 40. Together these results indicate no floor or ceiling effects for the 16-item PSWQ and for the 8-item PSWQ-A. For the PSWQ-3, scores ranged from 3 to 15 in the combined sample, with 16.7% achieving the lowest score of 3, and 6% achieved the highest possible score of 15. These results indicate a presence of floor effects, but not ceiling effects for the PSWQ-3.

Construct Validity

For adults with a diagnosis of GAD, all three versions of the PSWQ (PSWQ, PSWQ-A, and PSWQ-3) were found to moderate correlations with DASS-stress, IUS-12 and MCQ-30 negative metacognitive beliefs subscale, in line with hypotheses. The remaining questionnaires demonstrated a significant but low correlation with all three versions of the PSWQ, again in line with hypotheses. See Table 4 for correlations.

Internal Consistency

For adults with a diagnosis of GAD, the Cronbach's alpha for the 16-item PSWQ (α = 0.86) and the 8-item PSWQ-A (α = 0.80) was good. The Cronbach's alpha for the PSWQ-3 (α = 0.74) was adequate.

Test-retest Reliability

The ICC for all three versions indicated moderate reliability over a 12-week period: PSWQ (ICC r (19) = 0.66, $p < .01$), PSWQ-A [ICC (19) = 0.66, $p < .01$] and PSWQ-3 [ICC (19) = 0.56, $p = .04$].

Criterion Validity

The 16-item PSWQ was correlated using a spilt-half sample, and the strength was excellent for the PSWQ-A (r (140) = 0.95, $p < .001$) and good for the PSWQ-3 (r (140) = 0.89 $p < .001$). Criterion validity was also assessed using ROC curve analysis. Table 5 summarises the cut-off criteria and associated test performance indicators for the comparison of participants with GAD versus those without a mental health condition (i.e., non-clinical participants) for all three versions of the PSWQ. Of note, the AUC for the 16-item PSWQ, the PSWQ-A, and PSWQ-3 were all close to 1.0 (i.e., 0.99 − 0.98), indicating high accuracy for each of the respective cut-off scores.

Discussion

The current study compared the psychometric properties of the 16-item PSWQ, 8-item PSWQ-A and 3-item PSWQ-3 in a clinical sample of adults with GAD. Findings

Table 5 Receiver operating characteristic (ROC) curve analysis comparing adults with GAD ($n = 140$) to non-clinical participants ($n = 78$)

	PSWQ	PSWQ-A	PSWQ-3
Cut-Off Score	> 50	> 22	> 7
Area Under the Curve (AUC) (95% CI)	0.99 ** (0.97–0.99)	0.98** (0.95 − 0.99)	0.99 ** (0.96 − 0.99)
Sensitivity (95% CI)	97.14 (92.8–99.2)	96.43 (91.9–98.8)	94.29 (89.1–97.5)
Specificity (95% CI)	94.74 (87.1–98.5)	90.79 (81.9–96.2)	93.42 (85.3–97.8)
Positive Predictive Value (95% CI)	97.12 (92.9–98.9)	95.16 (90.5–97.5)	96.48 (91.9–98.4)
Negative Predictive Value (95% CI)	94.75 (87.2–97.9)	93.23 (85.3–97.0)	89.95 (81.9–94.6)

Note. *PSWQ* Penn State Worry Questionnaire, *PSWQ-A* PSWQ-Abbreviated, *95% CI* 95% Confidence Interval

** $p < .001$

demonstrated that all three versions of the PSWQ possess good psychometric properties across a range of indicators. Despite their brevity, the 8-item PSWQ-A and PSWQ-3 uphold their psychometric properties in comparison to the longer form and can be endorsed for use in both clinical and research settings for adults with GAD.

Initially six different confirmatory factor models were fit to the data, with four models incorporating 16-items from the PSWQ (i.e., unidimensional, bifactor, two factor, one factor with method effects), a unidimensional model for the 8-item PSWQ-A, and a unidimensional model for the PSWQ-3. The bifactor model appeared to fit the 16-item PSWQ better than the other PSWQ model iterations, though there was some room for improvement in absolute (i.e., χ^2 test) and incremental (TLI) fit indices. This finding endorses previous research that compared different versions of the 16-item PSWQ showing that the three-factor solution (i.e., one general factor and two method factors) provided the best fit to the data, mirroring current findings indicating that the longer form can be further improved, in a clinical GAD sample (Dear et al., 2011). Regarding bifactor indices from the current study, the general PSWQ factor represented the dominant source of variance (ω_H) in the total PSWQ score, with the *H* value indicating that the general factor was a well-defined latent variable (H). In addition, the ECV was greater than 0.70, suggesting that the general worry factor is sufficiently unidimensional to be treatment as a latent variable, thus using a total score for the PSWQ is recommended. Together, these results provide support for a strong general PSWQ factor, and unidimensionality (over multidimensionality) for the PSWQ. The unidimensional model also demonstrated the best fit the 8-item PSWQ-A across all absolute and incremental indices. This is perhaps unsurprising, as the PSWQ-A was developed by removing the negatively worded items. Regarding the PSWQ-3, comparable analyses could not be conducted on the PSWQ-3 because the model was saturated.

The psychometric properties of the 16-item PSWQ, 8-item PSWQ-A, and 3-item PSWQ-3 were largely comparable in their results and supported most hypotheses. Of note, all three versions showed moderate positive correlations with physiological stress/tension. This demonstrated construct validity as physiological tension and vigilance symptoms are part of the diagnostic criteria for GAD. All three measures also demonstrated moderate positive relationship with (1) negative metacognitive beliefs and (2) intolerance of uncertainty. These processes of are particular importance as they relate to two dominant models of GAD (Freeston, 2023). The metacognitive model proposes that excessive, pathological worry is primarily the result of negative metacognitive beliefs about worrying (i.e., that worry is uncontrollable and dangerous), in combination with positive beliefs about worry and subsequent ineffective mental control strategies (Wells, 2010). Whereas the intolerance of uncertainty model of GAD suggests that uncertainty is a natural trigger for worry, and therefore individuals who hold negative beliefs about uncertainty (i.e., intolerance of uncertainty) are more likely to experience excessive and difficult to control worry (Dugas et al., 1998; Hebert & Dugas, 2019). The hypotheses also predicted, in keeping with previous research (Kertz et al., 2014; Wells & Cartwright-Hatton, 2004), that less salient components of the metacognitive model, as well as distinct yet overlapping psychopathology constructs (i.e., depression and autonomic anxiety), would show positive relationships with the excessive worry. Together these findings are supported by the broader literature (Freeston, 2023), and highlight the importance for clinicians to target negative metacognitive beliefs about worry, as well as intolerance of uncertainty in psychological treatment, given that these processes were most strongly related to worry in the present study.

In terms of equivalence, all three versions demonstrated moderate test-retest reliability, excellent criterion validity, as well as the ability to distinguish adults with GAD from those with no mental health conditions in the non-clinical group. Criterion validity was assessed using a split-half sample analysis and demonstrated a strong correlation between the PSWQ and each of the two briefer versions (PSWQ-A and PSWQ-3). Though this result is not surprising, it strongly indicates that all three versions are measuring the same construct of excessive worry. One discrepancy between the three versions related to internal consistency, with the PSWQ-3 demonstrating adequate internal consistency, which was slightly lower than that for the PSWQ-A

and PSWQ. This may be partly accounted for by the PSWQ-3's substantial reduction in items, while still falling within an acceptable range (i.e., $r > 0.70$) to meet quality criteria for measurement properties set out by Terwee et al. (2007). All three versions showed a statistically significant difference between the clinical GAD group and the non-clinical group. The PSWQ and PSWQ-A demonstrated no floor or ceiling effects, however, the PSWQ-3 demonstrated a propensity for floor effects suggesting a lack of specificity. Though floor effects were found for the PSWQ-3, ROC curve analysis demonstrated that all three versions showed an AUC close to 1.0, indicating high accuracy for each of the respective cut-off scores that maximises both sensitivity and specificity to distinguish adults with GAD from non-clinical adults. It is noteworthy that despite the substantial reduction in items, the psychometric properties of the briefer versions are largely equivalent to the 16-item PSWQ.

The two briefer versions (i.e., PSWQ-A and PSWQ-3) appear to have relatively comparable psychometric properties to the original 16-item PSWQ in a GAD sample. Researchers and clinicians who need to measure pathological worry while accounting for time constraints may consider the two shorter versions of the PSWQ: the PSWQ-A and PSWQ-3. These briefer forms are particularly beneficial in clinical assessments where a broad range of symptoms are being evaluated, as they help mitigate client questionnaire fatigue. Both the PSWQ-A and PSWQ-3 demonstrate psychometric properties comparable to the full PSWQ, making them effective and efficient options for inclusion in assessment batteries. Another instance where time constraints may favour the use of a briefer version is in tracking weekly treatment progress for adults with GAD. The ultra-brief PSWQ-3 is particularly advantageous in this context due to its quick administration time and its resistance to floor effects within a GAD sample. Another consideration for researchers and clinicians to consider when using the 16-item PSWQ is that not only do the negatively worded items appear to increase the cognitive load when completing the 16-item PSWQ, they can make real-time scoring on pen and paper forms difficult for clinicians. As without a reverse-score template, clinicians are not able to quickly inspect whether the general pattern of worry is decreasing (or increasing) over the course of treatment. One way to overcome this difficulty for clinicians, would be to automate scoring for the 16-item PSWQ through a secure survey platform (i.e., REDCap). However, for clinicians who prefer to stick to pen and paper forms, the briefer versions of the PSWQ provide a readily available solution, that is psychometrically comparable. Thus, not only do the briefer versions avoid the methodological problems related to the reverse scored items on the 16-item PSWQ, but they also increase the speed of completion for participants or consumers in clinical practice.

It is important to highlight potential limitations of the study. First, the PSWQ-A and PSWQ-3 were not administered separately from the 16-item PSWQ, similarly to the methodology used in Wuthrich et al. (2014). Though this avoids potential problems with repeated measure effects, it prevents direct comparison of the measure's performance outside of the context of the full version. This limitation may also mean that there is an overestimation of the correlation between the versions, as the items on the 8-item and 3-item form are counted in both sides of the correlation (Smith et al., 2000). Item position has also been shown to impact questionnaire results (Podsakoff et al., 2012), which may have influenced the current findings. In addition, treatment sensitivity was not examined in the current study, previous research has found that the PSWQ (Dear et al., 2011) and the PSWQ-3 (Berle et al., 2011) were sensitive to evidence-based treatment with significant reductions following CBT, however, the PSWQ-A is yet to be assessed in adults with GAD. Another limitation of the current study was the inability to report on ethnicity and race for both clinical and non-clinical groups. The current sample had a large proportion of missing data for ethnicity in the GAD sample, with only 18.6% of the clinical group reporting their ethnicity. Future research should endeavour to capture ethnicity and race in samples, so that cultural differences can be explored and understood in the reporting of the short and longer forms of the PSWQ. In addition, a relatively small sub-sample of participants with GAD to assess test-retest reliability. This is a methodological limitation as quality criteria set out by Terwee et al. (2007) suggests that a sample size of at least 50 participants is required to meet adequate quality criteria for assessing psychometric properties of a questionnaire. Thus, the current study provides preliminary evidence that all three versions of the PSWQ demonstrate adequate test-retest reliability in adults with GAD, and future research should attempt to replicate this in a larger sample.

Overall, the shorter versions (PSWQ-A and PSWQ-3) performed with psychometric equivalence to the full 16-item PSWQ in a sample of individuals with GAD. Therefore, clinicians and researchers may prefer to utilise these self-report instruments, since they are both quicker to complete and score, and avoid scoring and psychometric issues caused by reverse-scored items, thereby potentially easing questionnaire burnout in the research space as well as facilitating real-time discussions in clinical practice.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

References

Abbott, M. J. (2007). Enhancing outcomes in the treatment of generalised anxiety disorder. *Innovations and advances in cognitive Behaviour Therapy*. Australian Academic.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berle, D., Starcevic, V., Moses, K., Hannan, A., Milicevic, D., & Sammut, P. (2011). Preliminary validation of an ultra-brief version of the Penn state worry questionnaire. *Clinical Psychology & Psychotherapy*, *18*(4), 339–346. https://doi.org/10.1002/cpp.724

Boehnke, K., Schwartz, S., Stromberg, C., & Sagiv, L. (1998). The structure and Dynamics of worry: Theory, measurement, and cross-national replications. *Journal of Personality*, *66*(5), 745–782. https://doi.org/10.1111/1467-6494.00031

Borkovec, T. D. (1994). The nature, functions, and origins of worry. *Worrying: Perspectives on theory, assessment and treatment* (pp. 5–33). Wiley.

Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (ROC) Area under the curve (AUC): A diagnostic measure for evaluating the Accuracy of predictors of Education outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734

Brown, T. A. (2003). Confirmatory factor analysis of the Penn state worry questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*, *41*(12), 1411–1426. https://doi.org/10.1016/S0005-7967(03)00059-7

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*(1), 33–37. https://doi.org/10.1016/0005-7967(92)90093-V

Brown, T. A., Barlow, D. H., & DiNardo, P. A. (1994). *Anxiety disorders interview schedule for DSM-IV: Adult Version. Client interview schedule*. Oxford University Press.

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*(1), 79–89. https://doi.org/10.1016/S0005-7967(96)00068-X

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Chyung, S. Y. Y., Barkin, J. R., & Shamsy, J. A. (2018). Evidence-based Survey Design: The Use of negatively worded items in surveys. *Performance Improvement*, *57*(3), 16–25. https://doi.org/10.1002/pfi.21749

Constans, J. I., Barbee, J. G., Townsend, M. H., & Leffler, H. (2002). Stability of worry content in GAD patients: A descriptive study. *Journal of Anxiety Disorders*, *16*(3), 311–319. https://doi.org/10.1016/S0887-6185(02)00102-0

Crittendon, J., & Hopko, D. R. (2006). Assessing worry in older and younger adults: Psychometric properties of an abbreviated Penn state worry questionnaire (PSWQ-A). *Journal of Anxiety Disorders*, *20*(8), 1036–1054. https://doi.org/10.1016/j.janxdis.2005.11.006

Dear, B. F., Titov, N., Sunderland, M., McMillan, D., Anderson, T., Lorian, C., & Robinson, E. (2011). Psychometric comparison of the generalized anxiety disorder Scale-7 and the Penn State worry questionnaire for measuring response during treatment of generalised anxiety disorder. *Cognitive Behaviour Therapy*, *40*(3), 216–227. https://doi.org/10.1080/16506073.2011.582138

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models.

Dugas, M. J., Freeston, M. H., Ladouceur, R., Rhéaume, J., Provencher, M., & Boisvert, J. M. (1998). Worry themes in primary GAD, secondary GAD, and other anxiety disorders. *Journal of Anxiety Disorders*, *12*(3), 253–261. https://doi.org/10.1016/S0887-6185(98)00013-9

Flora, D. B., & Flake, J. K. (2017). The purpose and Practice of Exploratory and Confirmatory Factor Analysis in Psychological Research: Decisions for Scale Development and Validation. *Canadian Journal of Behavioural Science*, *49*(2), 78–88. https://doi.org/10.1037/cbs0000069

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its Associated Cutoff Point. *Biometrical Journal*, *47*(4), 458–472. https://doi.org/10.1002/bimj.200410135

Freeston, M. H. (2023). What if we have too many models of worry and GAD? *Behavioural and Cognitive Psychotherapy*, *51*(6), 559–578. https://doi.org/10.1017/S1352465822000649

Freeston, M. H., Rhaume, J., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences*, *17*(6), 791–802.

Fresco, D. M., Heimberg, R. G., Mennin, D. S., & Turk, C. L. (2002). Confirmatory factor analysis of the Penn state worry questionnaire. *Behaviour Research and Therapy*, *40*(3), 313–323. https://doi.org/10.1016/S0005-7967(00)00113-3

Hancock, G., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and Future*, 195–216.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Hazlett-Stevens, H., Ullman, J. B., & Craske, M. G. (2004). Factor structure of the Penn State worry questionnaire: Examination of a method factor. *Assessment*, *11*(4), 361–370. https://doi.org/10.1177/1073191104269872

Hebert, E. A., & Dugas, M. J. (2019). Behavioral experiments for intolerance of uncertainty: Challenging the unknown in the treatment of generalized anxiety disorder. *Cognitive and Behavioral Practice*, *26*(2), 421–436. https://doi.org/10.1016/j.cbpra.2018.07.007

Hopko, D. R., Reas, D. L., Beck, J. G., Stanley, M. A., Wetherell, J. L., Novy, D. M., & Averill, P. M. (2003). Assessing worry in older adults: Confirmatory factor analysis of the Penn State worry questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment*, *15*(2), 173–183. https://doi.org/10.1037/1040-3590.15.2.173

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modelling: A Multidisciplinary Journal, 6, 1–55. https://doi.org/10.1080/10705519909540118

Huntley, C., Young, B., Smith, C. T., Jha, V., & Fisher, P. (2020). Assessing metacognitive beliefs in test anxiety: Psychometric

properties of the metacognitions questionnaire, 30 (MCQ-30) among university students. *Current Psychology*, *41*, 1425–1433.
Kertz, S. J., Lee, J., & Björgvinsson, T. (2014). Psychometric properties of abbreviated and ultra-brief versions of the Penn state worry questionnaire. *Psychological Assessment*, *26*(4), 1146–1154. https://doi.org/10.1037/a0037251
Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236
Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
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*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U
McEvoy, P. M., Salmon, K., Hyett, M., Jose, P., Gutenbrunner, P., Bryson, C., K., & Dewhirst, M. (2019a). Repetitive negative thinking as a transdiagnostic predictor of depression and anxiety symptoms in adolescents -. *Assessment*, *26*, 324–335.
McEvoy, P. M., Hyett, M. P., Shihata, S., Price, J. E.,& Strachan, L. (2019b). The impact of methodological and measurement factors on transdiagnostic associations with intolerance of uncertainty: A meta-analysis. Clinical Psychology Review, 73, 101778. https://doi.org/10.1016/j.cpr.2019.1017
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some Intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.
Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6
Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & De Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006
Molina, S., Borkovec, T. D., Peasley, C., & Person, D. (1998). Content analysis of Worrisome streams of consciousness in anxious and dysphoric participants. *Cognitive Therapy and Research*, *22*, 109–123.
Nordahl, H., Anyan, F., & Hjemdal, O. (2023). Prospective relations between dysfunctional metacognitive beliefs, metacognitive strategies, and anxiety: Results from a four-Wave Longitudinal Mediation Model. *Behavior Therapy*, *54*(5), 765–776. https://doi.org/10.1016/j.beth.2023.02.003
Page, A. C., Hooke, G. R., & Morrison, D. L. (2007). Psychometric properties of the Depression anxiety stress scales (DASS) in depressed clinical samples. *British Journal of Clinical Psychology*, *46*(3), 283–297. https://doi.org/10.1348/014466506X158996
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and recommendations on how to control it. *Annual Review of Psychology*, *63*(63, 2012), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452
Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the Presence of Multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437
Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). The Reverse of Social anxiety is not always the Opposite: The reverse-scored items of the Social Interaction anxiety scale do not Belong. *Behavior Therapy*, *38*(2), 192–206. https://doi.org/10.1016/j.beth.2006.08.001
Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249
Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113–130. https://doi.org/10.1080/02602930802618344
Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102
Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th ed.). Boston, MA Pearson.
Terwee, C. B., Bot, S. D. M., De Boer, M. R., Van Der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & De Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
Tucker, L. R., & Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*, *38*(1), 1–10. doi:10.1007/BF02291170
Wells, A. (2010). Metacognitive theory and therapy for worry and generalized anxiety disorder: Review and status. *Journal of Experimental Psychopathology*, *1*(1), jep007910. https://doi.org/10.5127/jep.007910
Wells, A., & Cartwright-Hatton, S. (2004). A short form of the metacognitions questionnaire: Properties of the MCQ-30. *Behaviour Research and Therapy*, *42*(4), 385–396. https://doi.org/10.1016/S0005-7967(03)00147-5
Wilson, E. J., Stapinski, L., Dueber, D. M., Rapee, R. M., Burton, A. L., & Abbott, M. J. (2020). Psychometric properties of the intolerance of uncertainty Scale-12 in generalized anxiety disorder: Assessment of factor structure, measurement properties and clinical utility. *Journal of Anxiety Disorders*, *76*, 102309. https://doi.org/10.1016/j.janxdis.2020.102309
Wuthrich, V. M., Johnco, C., & Knight, A. (2014). Comparison of the Penn State worry questionnaire (PSWQ) and abbreviated version (PSWQ-A) in a clinical and non-clinical population of older adults. *Journal of Anxiety Disorders*, *28*(7), 657–663. https://doi.org/10.1016/j.janxdis.2014.07.005
Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for Pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204–221. https://doi.org/10.1093/jpepsy/jst062