



Published in final edited form as:

*Qual Life Res.* 2013 December ; 22(10): . doi:10.1007/s11136-013-0395-8.

## Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO)

Taehwan Park<sup>1</sup>, Maryanne Reilly-Spong<sup>1</sup>, and Cynthia R Gross<sup>1</sup>

Cynthia R Gross: gross002@umn.edu

<sup>1</sup>University of Minnesota, Minneapolis, MN USA

### Abstract

**Purpose**—Mindfulness has emerged as an important health concept based on evidence that mindfulness interventions reduce symptoms and improve health-related quality of life. The objectives of this study were to systematically assess and compare the properties of instruments to measure self-reported mindfulness.

**Methods**—Ovid Medline<sup>®</sup>, CINAHL<sup>®</sup>, and PsycINFO<sup>®</sup> were searched through May 2012, and articles were selected if their primary purpose was development or evaluation of the measurement properties (validity, reliability, responsiveness) of a self-report mindfulness scale. Two reviewers independently evaluated the methodological quality of the selected studies using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist. Discrepancies were discussed with a third reviewer, and scored by consensus. Finally, a level of evidence approach was used to synthesize results and study quality.

**Results**—Our search strategy identified a total of 2,588 articles. Forty-six articles, reporting 79 unique studies, met inclusion criteria. Ten instruments quantifying mindfulness as a unidimensional scale (n=5) or as a set of 2 to 5 subscales (n=5) were reviewed. The Mindful Attention Awareness Scale (MAAS) was evaluated by the most studies (n=27), and had positive overall quality ratings for most of the psychometric properties reviewed. The Five Facet Mindfulness Questionnaire (FFMQ) received the highest possible rating (“consistent findings in multiple studies of good methodological quality”) for two properties, internal consistency and construct validation by hypothesis testing. However, none of the instruments had sufficient evidence of content validity. Comprehensiveness of construct coverage had not been assessed; qualitative methods to confirm understanding and relevance were absent. In addition, estimates of test-retest reliability, responsiveness, or measurement error to guide users in protocol development or interpretation of scores were lacking.

**Conclusions**—Current mindfulness scales have important conceptual differences, and none can be strongly recommended based solely on superior psychometric properties. Important limitations in the field are the absence of qualitative evaluations and accepted external referents to support construct validity. Investigators need to proceed cautiously before optimizing any mindfulness intervention based on the existing scales.

### Keywords

Mindfulness; Instruments; Psychometrics; Measurement properties; COSMIN checklist; Systematic review

## Introduction

Mindfulness has emerged as an important concept in health and outcomes research, driven by a rapidly growing body of evidence that mindfulness training reduces symptoms and improves quality of life. Mindfulness training is the basis for widely accepted interventions in psychosomatic medicine and psychology [1]. Recent reviews have summarized evidence of the efficacy of these mindfulness interventions for persons with cancer [2], chronic medical conditions [3], and psychological disorders [4]. Clinical efficacy and durability have been shown for depression relapse prevention, anxiety reduction, and insomnia [4-7]. Recent meta-analyses estimated small to medium-sized treatment effects for the impact of mindfulness training on symptoms of stress, anxiety and depression [4,7,8]. Clinical trials of mindfulness training with health providers and community samples demonstrate significant improvements in stress management and enhanced well-being [9,10]. Mindfulness training has also been shown to improve biomarkers of glycemic control in diabetes [11,12], enhance immune response [13], and accelerate skin healing in psoriasis [14].

Mindfulness can be a dynamically changing state, a trait that differs between persons, and a skill that can be enhanced through training [15]. Drawing upon sources in Buddhist psychology, mindfulness has been described as arising from the intentional deployment of a triad of intertwined “behaviors of the mind” - attention, awareness and attachment; and defined as “the active maximizing of the breadth and clarity of awareness” [16]. Adapting mindfulness for use in health interventions for patients with cancer, Susan Bauer-Wu explains mindfulness as “Our capacity to intentionally bring awareness to present-moment experience with an attitude of openness and curiosity. It is being awake to the fullness of our lives right now, through engaging the five senses and noticing the changing landscapes of our minds without holding on to or pushing away from any of it” [17]. In preparing this review, we were guided by the two-part model of mindfulness proposed by Bishop and colleagues following a series of discussions among an interdisciplinary group of researchers [18]. This consensus model of mindfulness encompasses two components, attention and acceptance. The attention component pertains to maintaining awareness of present moment experience, and the acceptance component relates to the quality of relationship to experience (e.g., attitudes of openness and curiosity). This two-part conceptualization of mindfulness has been widely cited, and attention and acceptance are common elements across most definitions used in the construction of self-reports [19].

Although the mechanisms responsible for the health benefits of mindfulness are not known, clinical, experimental and brain imaging studies suggest increased symptom awareness, reduced emotional arousal, and greater engagement in health-promoting behaviors are involved [20,21]. Measuring mindfulness is important for research aimed at understanding its role in helping people to deal with emotional and physical health problems, and to guide refinements of mindfulness interventions to optimize health benefits [22,23]. It is appropriate to evaluate measures of mindfulness using a framework designed for patient reported outcomes (PROs), as large numbers of patients are being asked to complete mindfulness self-assessments in the context of health care and outcomes research [2,3,8]. Other personal factors highly salient to health maintenance and disease prevention which are not conventionally considered outcomes, such as self-efficacy, self-esteem, and perceived social support, have been included in popular frameworks for PROs and self-report instruments to measure these health-related factors have been developed using PRO guidelines (<http://www.nihpromis.org/>).

Numerous studies have evaluated self-report instruments to quantify mindfulness [23-25], however a comprehensive, systematic review of these instruments has not been conducted using a level of evidence approach. The level of evidence approach relies upon

systematically ranking studies by the rigor of their methods so that final recommendations reflect results from the most methodologically sound studies. Conclusions from a level of evidence approach consider both the consistency of findings across studies and the rigor of those studies. The strongest evidence derives from consistent findings from multiple studies judged to have good or excellent methods. The aim of this study is to critically appraise and summarize the quality of the measurement properties of all published self-report mindfulness instruments using a level of evidence approach and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines [26]. COSMIN uses a taxonomy of measurement properties selected for relevance to health-related PROs based on the consensus of an international team of experts in health outcomes research. COSMIN includes uniform definitions and standards for the evaluation of methodological quality of studies to be reviewed, and has been used in more than a dozen systematic reviews published in peer-reviewed journals.

## Methods

### Search strategy

The electronic databases Ovid Medline<sup>®</sup> (1949 through May 2012), CINAHL<sup>®</sup> (1981 through May 2012), and PsycINFO<sup>®</sup> (1806 through May 2012) were searched using mindfulness index terms in combination with psychometric terms as described in Appendix 1. A manual search of the references of the included studies was conducted to supplement the electronic search. The search was limited to articles published in the English language.

### Selection criteria

Articles were selected if their primary purpose was to develop or evaluate the measurement properties of an original version of a mindfulness instrument. The instrument had to quantify mindfulness, and be developed for self-administration by adults. Instruments that were program-specific were excluded as were instruments that did not measure mindfulness *per se*. Therefore, instruments to measure mindful eating [27], mindful coping [28], meditation experience [29], mindfulness practice [30], self-compassion [31], and mindfulness-based relapse prevention adherence and competence (MBRP-AC) [32] were excluded. Articles were excluded if they were not full-text, original articles (e.g., reviews, commentaries, or dissertations) or if they were designed to create a brief, translated, or adolescent/child's version of another mindfulness scale. Articles about mindfulness instruments originally developed in any language other than English were initially excluded, however, after review of the collected articles, an exception to this rule was made to include the Freiburg Mindfulness Inventory because of its importance to the field as the first insight meditation-inspired self-report measure of mindfulness. Articles were also excluded if the primary aim was to test the efficacy of a mindfulness intervention. The decision to exclude efficacy trials was based on the recommendation for the conduct of systematic reviews from the text by De Vet et al. [33]. These authors note that efficacy studies generally provide only indirect evidence on the measurement properties of an instrument and this evidence is often difficult to interpret. Efficacy trials have been the focus of a growing number of meta-analyses.

One reviewer (T.P.) conducted the initial screening of titles and abstracts for all articles retrieved by the literature search, and identified candidate articles. Two reviewers (T.P. and C.R.G.) assessed the full text of the candidate articles, and jointly made decisions regarding article inclusion.

### Measurement properties

The COSMIN taxonomy groups psychometric properties into three domains: reliability, validity, and responsiveness [34]. Reliability, the degree to which an instrument is free from

measurement error, includes three properties: internal consistency, measurement error, and reliability. Internal consistency is the degree of the interrelatedness among the items in an instrument, and is typically assessed by Cronbach's alpha. Measurement error is the systematic and random error that is not attributed to true changes in the underlying construct, and it is adequate if the smallest detectable change (SDC) on the instrument is less than the minimal important change (MIC) [35]. Reliability is the proportion of the total variance reflecting true differences between persons, assessed by intraclass correlation coefficients (ICCs), Cohen's Kappa, or test-retest correlations. Validity is the extent to which an instrument measures the construct it purports to measure. COSMIN groups three properties in the validity domain: content validity, construct validity, and criterion validity. Content validity includes face validity, comprehensiveness, and relevance of the items in an instrument for its target population and purpose. Structural validity, hypothesis testing, and cross-cultural validity are aspects of construct validity. Structural validity is the evidence to support the dimensionality of an instrument, and hypothesis testing is the degree to which relationships between an instrument and other measures conform to expectations, including differences between known groups. Relationships are often assessed by the Pearson correlation coefficient ( $r$ ). Criterion validity, the extent that an instrument correlates with an accepted "gold standard" was not applicable for this review, because there is no gold standard for mindfulness. Cross-cultural validity, the extent that items of a translated or adapted version of an instrument perform as items on the original perform, was also not evaluated for this review. The final domain is responsiveness, the ability of an instrument to detect change in the underlying construct. To assess responsiveness, investigators pose hypotheses about expected correlations between the change score on the target instrument and change scores on other instruments for the same or other constructs. This is essentially validity in a longitudinal context. Responsiveness is not assessed by treatment effect size in COSMIN. This approach is consistent with the approach of Brown & Ryan, who noted that "the present study was not designed to test the efficacy of the intervention *per se*, but rather to examine whether mindfulness and changes in it were related to well-being outcomes and changes in them" [15].

### The COSMIN checklist and study quality assessment

An evidence-based approach requires that results be relied upon only when they are produced by methodologically sound studies. The COSMIN checklist contains 98 items to assess whether a study of measurement properties meets quality standards [36]. Study quality is determined separately for each measurement property, using 5 to 18 items each rated as *poor*, *fair*, *good*, or *excellent*. The final quality rating for a property is the lowest rating of any item pertinent to that property (worst rating counts) [36].

Two reviewers (T.P. and C.R.G.) independently extracted data from the selected articles and evaluated methodological quality using the COSMIN checklist as a guide. Discrepancies were discussed with a third reviewer (M.R.-S.) to reach consensus. Where a single article presented multiple studies, each study was separately evaluated and rated for every measurement property it addressed.

### Best evidence synthesis

Study findings were rated as "positive," "negative," or "indeterminate" for each measurement property based on criteria proposed by Terwee et al (Table 1) [37]. Summaries for each instrument were prepared showing how many studies of *excellent*, *good*, *fair*, or *poor* quality provided positive, negative, or indeterminate results by property. Overall ratings were then synthesized for each instrument across studies using a level of evidence approach that considered the number and methodological quality of the studies, and the

consistency of their findings (Table 2) [38]. Findings from *poor* quality studies received no weight in the final synthesis.

## Results

The study selection process is presented in Figure 1. A total of 2,588 unique articles were identified using the search strategy; of these, 146 articles were selected based on their title and abstract. For further assessment, the full text of these articles was examined resulting in the exclusion of 67 articles. As shown in Figure 1, most of these were excluded because evaluation of psychometric properties was not the primary focus of the article. Another 33 articles were then excluded because they addressed a translated, short or modified version of the original instruments, leaving a total of 46 articles for inclusion in our review. These 46 articles contained 79 separate studies, and evaluated 10 different mindfulness instruments. Several studies evaluated multiple instruments. Table 3 shows the characteristics of the included studies, Table 4 presents the characteristics of the included instruments, and Table 5 shows the COSMIN ratings of the methodological quality of these studies, by measurement property. No selected article was completely excluded for poor methodological quality. Our synthesis of the results and level of evidence for the properties of each mindfulness instrument is presented in Table 6. Results for each instrument are summarized below. In these summaries we use the following conventions for describing correlations: correlations are considered strong if  $|r|$  is between 0.7 and 1.0, moderate if  $0.4 < |r| < 0.7$ , and weak if  $0 < |r| < 0.4$ . Results from studies of *poor* methodological quality are not included in these summaries.

### Mindfulness Attention Awareness Scale (MAAS)

The MAAS was the first widely-disseminated measure of mindfulness. It was designed to measure mindfulness as present-centered attention-awareness in everyday experience, a state which varies within and between persons, and an attribute that may be cultivated with practice [15]. This instrument focused on the absence of attention to and awareness of present experience, and was designed to operationalize mindfulness as a single construct. This instrument was intended to be generic, and applicable to persons regardless of experience with meditation. Sample items are shown in Table 4. Most studies confirmed a 1-factor structure for the MAAS [15,39-42]. One study found that some items in the MAAS did not function well as indicators for a single latent construct [43]. There was support for the internal consistency of the MAAS (Cronbach alphas ranging from 0.78 to 0.92) and evidence of test-retest reliability (ICC = 0.81). Correlations between the MAAS and other mindfulness instruments, such as the FMI, CAMS-R, SMQ, KIMS, and MMS, were weak to moderate ( $r$ 's = 0.14 to 0.51) [15,44,45]. Consistent with expectations for construct validity, MAAS scores were positively correlated with measures of openness, internal state awareness, positive and pleasant affect, and well-being, and negatively correlated with neuroticism, anxiety, stress, and rumination [15,39,44-47]. MAAS scores were higher for meditators compared to non-meditators [43], but there was no significant difference between novice meditators and non-meditators [40]. Several studies [15,48,49] have compared the MAAS to results on performance-based tasks (e.g., cognitive tests of attention, inhibition) with mixed results.

### Kentucky Inventory of Mindfulness Skills (KIMS)

The KIMS was designed to assess the tendency to be mindful in daily life in areas corresponding to the skills taught in mindfulness interventions, particularly Dialectical Behavior Therapy [50]. The KIMS consists of 39 items grouped into four subscales: Observe, Describe, Act with Awareness, and Accept without Judgment. The Observe subscale reflects the skill of observing or paying attention to internal (bodily sensations,

thoughts and emotions) and external phenomena. The Describe subscale refers to a tendency or ability to put sensations, perceptions, thoughts, feelings, emotions, or experiences into words. The Act with Awareness subscale reflects the ability to focus undivided attention on the present. The Accept without Judgment subscale includes both the act of making judgments and common examples of self-criticism. The 4-factor structure of the KIMS was supported by exploratory factor analysis (EFA); 43% of the variance was accounted for by the 4-factors [50]. Although nearly adequate fit was shown in confirmatory factor analysis (CFA), the analyses used a somewhat controversial “parceling approach” to overcome CFA sample size limitations, and others were unable to replicate the 4-factor solution by EFA [41]. The KIMS (global and subscales) had evidence of internal consistency (Cronbach alphas ranging from 0.72 to 0.97), and test-retest reliability was adequate ( $r$ 's ranging from 0.81 to 0.86) for all but the Observe subscale ( $r = 0.65$ ) [50]. The construct validity of the KIMS global score was supported by moderate correlations ( $r$ 's ranging from 0.51 to 0.67) with the MAAS, FMI and CAMS-R and positive correlations with meditation experience [44]. Consistent with expectations for convergent and divergent validity, the global KIMS had positive correlations with openness, emotional intelligence, and self-compassion, and negative correlations with psychological symptoms, neuroticism, alexithymia, dissociation, and absent-mindedness [44]. KIMS subscales had different levels of evidence to support their construct validity. Accept without Judgment has consistently been found to be the most robust subscale, with most a priori relationships with health and quality of life measures confirmed [45,47,50-52]. There was also moderate evidence of the construct validity of the Act with Awareness subscale. Evidence to support the construct validity of the Describe subscale was limited, and relationships with the Observe subscale have been unpredictable. For example, the Observe subscale did not differ between adults with borderline personality disorder and normative student samples [50]. The developers acknowledged limitations in the content coverage of the KIMS, and concerns about integration of the subscales to provide a meaningful global score.

### Freiburg Mindfulness Inventory (FMI)

The FMI was originally developed and validated in German, and English translations of FMI items have been incorporated into more recently developed mindfulness instruments [44]. Buddhist psychology guided development of the FMI and its intended target audience was individuals with some knowledge about or familiarity with insight meditation. The FMI was designed to assess mindfulness as “attentional, unbiased observation of any phenomenon in order to perceive and to experience how it truly is, absent of emotional or intellectual distortion” [53]. The developers cited the hallmark of mindfulness as dispassionate, non-manipulative participant observation of ongoing mental states without conceptualizing or forming emotional reactions. EFA identified 4 factors for the FMI, however the structure was not stable across samples and items cross-loaded, which the authors interpreted as support for a single underlying factor [53]. This original 4-factor structure was only approximately replicated in a subsequent study [54], and these authors also favored interpreting the FMI as one general factor reflecting mindfulness. There was evidence to support the internal consistency of the global FMI (Cronbach alphas ranged from 0.80 to 0.94). The FMI had weak to moderate correlations with the MAAS, KIMS, CAMS-R, and SMQ ( $r$ 's = 0.31 to 0.60) [44]. As expected, the FMI was positively correlated with openness, self-compassion, and self-knowledge, and negatively correlated with psychological symptoms, neuroticism, difficulties in emotion regulation, alexithymia, dissociation, and distress [44,54]. However, there was an unexpected positive relationship between FMI scores and smoking/frequent binge-drinking among undergraduate college students, suggesting that the FMI may not be valid when completed by persons without some familiarity or experience with insight meditation [55]. This review pooled the findings from the original German version of the FMI [53,54] with those of its English translation

[44,55] because of the importance of the FMI as the first insight meditation-inspired self-report measure of mindfulness published.

### **Cognitive and Affective Mindfulness Scale-Revised (CAMS-R)**

The CAMS-R was designed to measure mindfulness in a brief, jargon-free, and conceptually comprehensive way, with the intention that it would be a generic measure appropriate regardless of meditation experience. Based on Kabat-Zinn's definition [56], "awareness that emerges through paying attention on purpose, in the present moment, and non-judgmentally to the unfolding of experience moment to moment," the authors conceptualized mindfulness as having four aspects: attention, present-focus, awareness, and acceptance/non-judgment [57]. Factor analyses provided moderate evidence of the predicted four aspects reflecting an overarching construct of mindfulness [57]. There was evidence of the internal consistency of the CAMS-R (Cronbach alphas ranging from 0.61 to 0.81). The CAMS-R had moderate correlations with other measures of mindfulness, including MAAS, FMI, KIMS, and SMQ ( $r$ 's = 0.51 to 0.67) [44,57]. Construct validity was supported by positive relationships with measures of adaptive regulation, openness, and well-being, and negative relationships with neuroticism, difficulties in emotion regulation, dissociation, and stagnant deliberation [44]. The CAMS-R, and not the original CAMS, was included in this review, because the developers determined that the CAMS was seriously flawed, and do not support its use [57].

### **Southampton Mindfulness Questionnaire (SMQ)**

The SMQ was designed to assess awareness of distressing thoughts and images defined as a concept consisting of four related constructs: awareness of cognitions as mental events in wider context, allowing attention to remain with difficult conditions, accepting such difficult thoughts and oneself without judging, and letting difficult cognitions pass without reactions such as rumination [58]. Although factor analysis suggested a single factor structure for the SMQ, a single-factor solution explained less than 50% of the variance [58]. There was evidence of the internal consistency of the SMQ (Cronbach alphas ranging from 0.82 to 0.89). Correlations between the SMQ and other measures of mindfulness varied from weak to moderate ( $r$ 's = 0.38 to 0.61) [44,58]. Consistent with expectations, the SMQ correlated positively with emotional intelligence and self-compassion, and negatively with neuroticism, difficulties in emotion regulation, alexithymia, dissociation, and negative affect [44,58]. SMQ scores were higher in meditators compared to non-meditators, and in non-clinical samples compared to patients with psychosis [58].

### **Five Facet Mindfulness Questionnaire (FFMQ)**

The FFMQ was derived from factor analysis of the combined item pool from five independently developed mindfulness instruments: MAAS, KIMS, FMI, CAMS-R, and SMQ [44]. The FFMQ has four facets similar to those of the KIMS (Observing, Describing, Acting with Awareness, and Nonjudging of inner experience) and one more facet comprised of items from the FMI and SMQ (Nonreactivity to inner experience). The authors found that the relationship between the facets and an overarching construct of mindfulness differed based on meditation experience, and that associations with symptoms and other constructs differed by facet. Therefore, they suggested use of the individual subscales may be preferred to use of the total FFMQ score. A 5-factor structure for the FFMQ was suggested by EFA [44] and confirmed by good or acceptable fit indexes in CFA using the same parceling approach for CFA employed in developing the KIMS [50,59]. A recent, standard item-level CFA supported the original 5-factor structure and an over-arching mindfulness factor [60]. Others have shown a modest fit for this structure [61], and hierarchical models that supported only four factors (all but Observe) as facets of an overarching mindfulness construct in student samples [44]. Internal consistency of the FFMQ is adequate with Cronbach alphas for the five subscales ranging from 0.67 to 0.93. Construct validity for the

global FFMQ and its subscales has been evidenced by positive correlations with openness, emotional intelligence, self-compassion, and well-being, and negative correlations with neuroticism, depression, anxiety, alexithymia, and dissociation [44,62-66]. Meditators scored higher on the FFMQ than non-meditating students, and meditation history was correlated with a total FFMQ score in meditating samples ( $r = 0.52$ ) [67]. The FFMQ Observe and Describe subscales were derived largely from the KIMS, and as with the KIMS, relationships with these subscales were less robust and predictable than those with other facets. For example, significant differences in Observe and Describe were not found between high- and low-worry groups [66]. There was little or no evidence for differential item functioning (DIF) between meditators and non-meditators matched for age [68], although the developers previously found that the structure of the FFMQ, particularly with respect to the Observe facet, differed between meditators and non-meditators [44].

### **Toronto Mindfulness Scale (TMS)**

The TMS was designed to assess mindfulness as a “quality maintained when attention is intentionally cultivated with an open, non-judgmental orientation to experience” [69]. The original TMS measures mindfulness as a state-like quality, and not as a trait. The administration of the TMS requires that a brief mindfulness exercise precede self-administration of the instrument, and the TMS items assess the quality of that experience. The TMS is composed of two subscales, Curiosity and Decentering, and a total TMS score is not reported. EFA suggested a 2-factor structure for the TMS, and this was supported by CFA [69]. The TMS had evidence of internal consistency with Cronbach alphas ranging from 0.86 to 0.91, and 0.85 to 0.87 for Curiosity and Decentering, respectively. Correlations for the Decentering subscale with most of the other measures of mindfulness, including MAAS, FMI, CAMS-R, SMQ, KIMS subscales, and FFMQ subscales ( $r$ 's = 0.20 to 0.74) were stronger than the correlations between the Curiosity subscale and these measures ( $r$ 's = 0.10 to 0.54) [70]. Curiosity and Decentering were positively correlated with absorption, awareness of surroundings, reflective self-awareness, and psychological mindedness. As hypothesized, only Curiosity was correlated with awareness of internal states and self-consciousness ( $r = 0.41$  and  $0.31$ ), and only Decentering was correlated with openness and cognitive failures ( $r = 0.23$  and  $-0.16$ ) [69]. Curiosity and Decentering scores were higher in meditators than non-meditators, and scores for the Decentering subscale were shown to increase with meditation experience [70]. Changes in Decentering were associated with changes in symptoms and stress [69].

### **Experiences Questionnaire (EQ)**

The EQ was designed to measure decentering, a construct described as the ability to adopt a wider perspective where one's thoughts are viewed as separate from oneself, and not necessarily an objective reflection of reality [71]. Decentering is posited to be a major outcome of mindfulness-based cognitive therapy and a mechanism that enables patients to be resilient to depressive thoughts. The authors did not view decentering as synonymous with mindfulness, but closely related or a component of mindfulness. The EQ was originally designed to have items reflecting decentering and rumination; however, the structure was determined to be unifactorial for the construct of decentering [71]. The EQ had evidence of internal consistency (Cronbach alphas ranging from 0.83 to 0.90), and construct validity was supported by positive correlations with cognitive appraisal ( $r = 0.25$ ), and negative correlations with experiential avoidance, brooding rumination, emotional suppression, current depression, and anxiety symptoms ( $|r$ 's = 0.31 to 0.49) [71]. Patients with depression had lower levels of decentering compared to healthy controls [71].



### Mindfulness/Mindlessness Scale (MMS)

The MMS was designed to assess mindfulness from a cognitive-information processing framework as active awareness of and engagement with the environment [72]. Its Western cognitive derivation distinguishes the MMS from the other measures presented in this review. The MMS is composed of four subscales: Novelty Seeking, Engagement, Novelty Producing, and Flexibility. The 4-factor structure has not been supported, and a 2-factor structure explaining 34% of the variance has been reported [72]. Evidence of internal consistency was positive for the MMS as a single scale with Cronbach alphas ranging from 0.81 to 0.86. Cronbach alphas for the MMS subscales ranged from 0.45 to 0.77. There was mixed evidence regarding the relationships between MMS items and measures of mood.

### Philadelphia Mindfulness Scale (PHLMS)

The PHLMS was designed to assess mindfulness defined as “the tendency to be highly aware of one’s internal and external experiences in the context of an accepting, nonjudgmental stance toward those experiences” [73]. This definition was operationalized as two constructs: Awareness - a behavioral tendency of continuously monitoring current experience, and Acceptance - a stance of experiencing events, including cognitions, without judgments and reactions such as interpretation, elaboration or avoidance. The subscales were shown to be uncorrelated, and use of a total PHLMS score is not recommended. A 2-factor structure for the PHLMS was supported by CFA [73]. Internal consistency was also supported with Cronbach alphas ranging from 0.75 to 0.86, and 0.75 to 0.91 for Awareness and Acceptance, respectively. Evidence of construct validity was mixed [73]. For example, the Awareness subscale was strongly correlated with the KIMS Observe subscale ( $r = 0.83$ ) and the Acceptance subscale was strongly correlated with the KIMS Accept without Judgment subscale ( $r = 0.79$ ) [73]. However, the correlation between the Awareness subscale and MAAS was weak ( $r = 0.21$ ) for student samples and moderate ( $r = 0.40$ ) for psychiatry outpatients. The correlation between the Acceptance subscale and MAAS was also weak ( $r = 0.32$ ) for the normative student samples. As expected, student samples scored higher on both PHLMS subscales than psychiatry outpatients, and students scored higher on the Acceptance subscale compared to the inpatients with eating disorders (EDs). However, Awareness scores were not significantly different between students and inpatients with EDs.

## Discussion

The purpose of this study was to systematically assess and compare the properties of instruments to measure self-reported mindfulness in adults. A comprehensive search strategy identified a total of 2,588 potentially relevant articles. Out of this pool, 46 articles reporting 79 unique studies, met the inclusion criteria for review. Ten instruments quantifying mindfulness as a unidimensional scale ( $n=5$ ) or as a set of 2 to 5 subscales ( $n=5$ ) were found. The COSMIN checklist was used to evaluate the methodological quality of each study for six properties in the COSMIN taxonomy: internal consistency, reliability, content validity, structural validity, hypothesis testing and responsiveness. We had initially planned to address measurement error, but no study evaluated this measurement property. The methodological quality of the studies included in this review was mostly *good* (66%) or *fair* (26%) across all properties. The majority of the studies were conducted with college undergraduates (48 out of 79 studies). No instrument had evidence to support content validity or adequacy of measurement error. The MAAS was the most frequently evaluated instrument followed by the KIMS. The MAAS was supported by positive evidence for internal consistency, reliability, construct validity by hypothesis testing, and responsiveness. The KIMS was supported by strong evidence for internal consistency, moderate evidence for construct validity by hypothesis testing, and limited evidence for reliability, but the other measurement properties were indeterminate or not available. The results shown in Table 6

provide limited guidance for instrument selection. The MAAS, KIMS, CAMS-R, FFMQ, TMS, EQ and PHLMS were found to have moderate or strong positive results for two or more properties; these measures may be preferred on psychometric grounds over the other instruments. Final instrument selection must consider other factors including the conceptual definition, completion time and target population. Moreover, as described below, there are areas where all the instruments are lacking; therefore caution is advised in using these results.

Descriptive critiques of mindfulness instruments have identified key problems including: 1) important differences in conceptual definitions of mindfulness; 2) no confirmation of respondent understanding of items; 3) absence of investigation of the potential discrepancies between self-reports and external referents (e.g., indicators of mindfulness experimentally tested or observed by others); and 4) conflation of the effects of learning the language of mindfulness or valuing mindfulness with actual increases in mindfulness *per se* [24,25]. To a great extent these problems are direct consequences of inadequate content validation. As documented by this systematic review, there was no engagement with members of a target population for item development and pre-testing. Cognitive interviews or focus groups to evaluate understanding and relevance to the target population or comprehensiveness of the items for the construct of mindfulness were not conducted. Neither was there any exploration of potential “response shift” in understanding of the construct of mindfulness following meditation training [74]. Moreover, the lack of diversity among the samples used in psychometric testing severely restricted the capacity of developers to detect potentially important differences among persons. It is unknown if items have very different semantic interpretations depending on the respondent’s characteristics, e.g., health status, age, race. Conceptual differences and lack of content validity were evidenced by weak to modest correlations among these measures of mindfulness, and among similarly titled mindfulness subscales. These gaps are consistent with a general lack of empirical studies comparing the psychometric performances of competing patient-reported outcome (PRO) measures within a complementary and alternative medicine (CAM) setting [75]. As no degree of superlative performance on other psychometric properties can compensate for poor content validity [76], none of the measures evaluated can be strongly recommended as a PRO at this time.

It is not clear as to which mindfulness instrument represents all the essential aspects of mindfulness. Some facets or dimensions of mindfulness may be more tractable to self-reports, and facets vary in their relationships with clinically relevant outcomes. “Summing up” purported facets of mindfulness as often done with the KIMS or FFMQ is likely to be problematic since some individuals who appear to possess higher levels of mindfulness could actually have a “toxic” combination of mental behaviors, such as being highly aware and very judgmental [16,25,77]. Although cogent arguments have been made for the utility of a brief, all-inclusive measure of mindfulness (e.g., CAMS-R) for clinical use, others have urged that instruments address specific sub-domains and be re-titled to better reflect their contents, and avoid having a multiplicity of instruments with very different content all claiming to measure mindfulness [25,78].

### Utility of Mindfulness Scales

There is a surprising lack of information to guide users of these instruments. Few instruments had information on test-retest reliability or responsiveness, and none provided evidence of the adequacy of measurement error or estimated a minimally important difference. Floor or ceiling effects, rates of missing data, average completion time, and skewness of distributions were mentioned rarely or not at all. For the instruments with subscales, additional guidance regarding whether or not subscales should or should not be

combined and how those scores should be labeled (e.g., total or global) and reported (e.g., mean or sum) would promote consistency and facilitate comparisons across studies.

### Use of COSMIN and Quality Criteria

The COSMIN checklist is a useful guide, but has shortcomings. First, benchmarks for sample size are not helpful for CFA, since they are based on number of items and not number of parameters to estimate, and do not account for approaches such as bundling items into parcels to overcome sample size limitations. We also found it necessary to better define thresholds for adequate fit of CFA. These are listed in the footnote to Table 1. Second, COSMIN weights reporting and handling of missing data very heavily. Studies that do not provide clear information about rates of missing data and explain how missing data were handled are rated as having no more than *fair* methodological quality on missing data items. We initially followed this guideline and it resulted in 71% of all studies receiving overall *fair* ratings for the property. We felt this under-represented the overall quality of the studies in this review. We therefore used a modified guideline to allow an overall rating of *good*, if the only flaw noted was inadequate reporting of missing data. These are the ratings shown in Table 5. Nevertheless, inadequate reporting and handling of missing data is problematic [79], and developers should be strongly encouraged to report rates of missing data and use robust methods for imputation.

It would be timely to update the quality criteria to assess measurement properties. These criteria derive from a 2007 paper by Terwee et al. [37], rely heavily upon classical test theory, and lack sufficient guidance for integrating findings from item response theory (IRT) into its quality ratings. For example, is an instrument to be down-graded for construct validity if several of its items are shown to have differential item functioning (DIF), and if so, how many items with DIF would result in a downgrade?

There are limitations to this study. First, only one investigator conducted the first stage review of the over 2,500 titles and abstracts identified by our search strategy. To verify the completeness of the initial selection, we relied upon our search of the references of the selected articles and investigation of citations for the selected instruments through the Web of Science. Second, selection bias may be introduced by including only studies published in English. We initially excluded all mindfulness instruments not developed in English, and then changed our criteria to include the German language FMI because of its importance to the field. We have noted where psychometric findings from the German and English versions have been pooled. Short forms of these mindfulness instruments were not included [80-82]. We cannot recommend use of any short form where the longer version lacks evidence of content validity; reducing the number of items will not overcome this serious flaw. Translated instruments were also not included. These instruments warrant a separate review to adequately address issues of meaning and cross-cultural validity.

In conclusion, self-reports of mindfulness have the potential to be an important means of assessing the mechanisms and outcomes of mindfulness-based therapies. There is a great need to establish the content validity of the extant measures of mindfulness using qualitative methods such as semi structured interviews and focus groups with novice and experienced meditators, diverse populations, and clinical populations with acute and chronic illnesses. Further explication of the construct of mindfulness, its facets and consequences, and pre-testing of items with diverse target populations to ensure comprehensiveness of content coverage, clarity, and relevance are needed. Items prone to bias from learning the language of mindfulness and recognizing its value should be eliminated. It is timely to devise external referents to validate these self-reports. External referents may take the form of neuropsychological or other performance tests, evaluations by third-parties, such as teachers, spouses, or other family members, biomarkers or imaging studies. Several of the brief,

royalty-free tests in the cognitive domain of the newly developed NIH Toolbox for Assessment of Neurological and Behavioral Function ([www.nihtoolbox.org](http://www.nihtoolbox.org)) may be useful external referents for mindfulness. Content validation should take precedence over efforts to optimize reliability and create short forms. Researchers using current mindfulness instruments are encouraged to report frequencies of skipped items to aid in identifying poor items for clinical samples, and estimate test-retest reliability, responsiveness, adequacy of measurement error, and minimally important differences. Use of mindfulness-based interventions continues to grow, with target populations and use of novel technologies for training rapidly expanding. Research to establish the best approaches for mindfulness training and target those most likely to benefit will be facilitated by valid and reliable self-reported mindfulness instruments.

## Acknowledgments

This study was supported in part by National Institutes of Health, National Institute of Diabetes and Digestive and Kidney diseases grant P01 DK 13083.

## APPENDIX 1. Full search strategy

Ovid Medline<sup>®</sup>, CINAHL<sup>®</sup>, and PsycINFO<sup>®</sup> were searched using the following search string:

1. mindful\* OR vipassana OR zen meditation OR insight mediation OR theravada OR Buddhist meditation
2. research measurement OR questionnaire\* OR scale\* OR instrument\* OR methods OR outcome assessment OR outcome measure OR psychometr\* OR reliab\* OR valid\* OR internal consistency OR (cronbach\* AND (alpha OR alphas)) OR (item AND (correlation\* OR selection\* OR reduction\*)) OR (intraclass AND correlation\*) OR interscale correlation\* OR agreement OR stability OR generaliza\* OR concordance OR variability OR kappa OR kappa's OR factor analysis OR factor analyses OR factor structure OR dimension OR subscale\* OR standard error of measurement OR test-retest OR (test AND retest) OR sensitiv\* OR responsive\* OR reproducib\* OR repeatab\* OR replica\* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR interpretab\* OR item response OR IRT OR Rasch OR differential item functioning OR ceiling effect\* OR floor effect\*
3. 1 and 2
4. limit to English language only.

## References

1. Baer, RA. Mindfulness-based treatment approaches: clinician's guide to evidence base and applications. Massachusetts: Academic Press; 2006.
2. Shennan C, Payne S, Fenlon D. What is the evidence for the use of mindfulness-based interventions in cancer care? A review. *Psycho-Oncology*. 2011; 20(7):681–697. [PubMed: 20690112]
3. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: A meta-analysis. *Journal of Psychosomatic Research*. 2010; 68(6):539–544. [PubMed: 20488270]
4. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2010; 78(2): 169–183. [PubMed: 20350028]

5. Gross CR, Kreitzer MJ, Thomas W, Reilly-Spong M, Cramer-Bornemann M, Nyman JA, et al. Mindfulness-based stress reduction for solid organ transplant recipients: A randomized controlled trial. *Alternative Therapies in Health and Medicine*. 2010; 16(5):30–38. [PubMed: 20882729]
6. Gross CR, Kreitzer MJ, Reilly-Spong M, Wall M, Winbush NY, Patterson R, et al. Mindfulness-based stress reduction versus pharmacotherapy for chronic primary insomnia: A randomized controlled clinical trial. *Explore: The Journal of Science & Healing*. 2011; 7(2):76–87.
7. Chiesa A, Serretti A. Mindfulness based cognitive therapy for psychiatric disorders: A systematic review and meta-analysis. *Psychiatry Research*. 2011; 187(3):441–453. [PubMed: 20846726]
8. Fjorback LO, Arendt M, Ornbol E, Fink P, Walach H. Mindfulness-based stress reduction and mindfulness-based cognitive therapy: A systematic review of randomized controlled trials. *Acta Psychiatrica Scandinavica*. 2011; 124(2):102–119. [PubMed: 21534932]
9. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: A review and meta-analysis. *Journal of Alternative and Complementary Medicine*. 2009; 15(5): 593–600.
10. Krasner MS, Epstein RM, Beckman H, Suchman AL, Chapman B, Mooney CJ, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. *Journal of the American Medical Association*. 2009; 302(12):1284–1293. [PubMed: 19773563]
11. Rosenzweig S, Reibel DK, Greeson JM, Edman JS, Jasser SA, McMearty KD, et al. Mindfulness-based stress reduction is associated with improved glycemic control in type 2 diabetes mellitus: A pilot study. *Alternative Therapies in Health and Medicine*. 2007; 13(5):36–38. [PubMed: 17900040]
12. Gregg JA, Callaghan GM, Hayes SC, Glenn-Lawson JL. Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*. 2007; 75(2):336–343. [PubMed: 17469891]
13. Davidson RJ, Kabat-Zinn J, Schumacher J, Rosenkranz M, Muller D, Santorelli SF, et al. Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine*. 2003; 65(4):564–570. [PubMed: 12883106]
14. Kabat-Zinn J, Wheeler E, Light T, Skillings A, Scharf MJ, Cropley TG, et al. Influence of a mindfulness meditation-based stress reduction intervention on rates of skin clearing in patients with moderate to severe psoriasis undergoing phototherapy (UVB) and photochemotherapy (PUVA). *Psychosomatic Medicine*. 1998; 60(5):625–632. [PubMed: 9773769]
15. Brown KW, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*. 2003; 84(4):822–848. [PubMed: 12703651]
16. Mikulas WL. Mindfulness: Significant common confusions. *Mindfulness*. 2011; 2(1):1–7.
17. Bauer-Wu, S. *Leaves falling gently: living fully with serious and life-limiting illness through mindfulness, compassion, and connectedness*. California: New Harbinger; 2011.
18. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, et al. Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*. 2004; 11(3):230–241.
19. Coffey KA, Hartman M, Fredrickson BL. Deconstructing mindfulness and constructing mental health: understanding mindfulness and its mechanisms of action. *Mindfulness*. 2010; 1(4):235–253.
20. Goldin PR, Gross JJ. Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion*. 2010; 10(1):83–91. [PubMed: 20141305]
21. Hölzel BK, Lazar SW, Gard T, Schuman-Olivier Z, Vago DR, Ott U. How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science*. 2011; 6(6):537–559.
22. Chiesa A, Malinowski P. Mindfulness-based approaches: Are they all the same? *Journal of Clinical Psychology*. 2011; 67(4):404–424. [PubMed: 21254062]
23. Sauer S, Lynch S, Walach H, Kohls N. Dialectics of mindfulness: Implications for western medicine. *Philosophy, Ethics, and Humanities in Medicine*. 2011; 6(10):1–7.
24. Grossman P. On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research*. 2008; 64(4):405–408. [PubMed: 18374739]

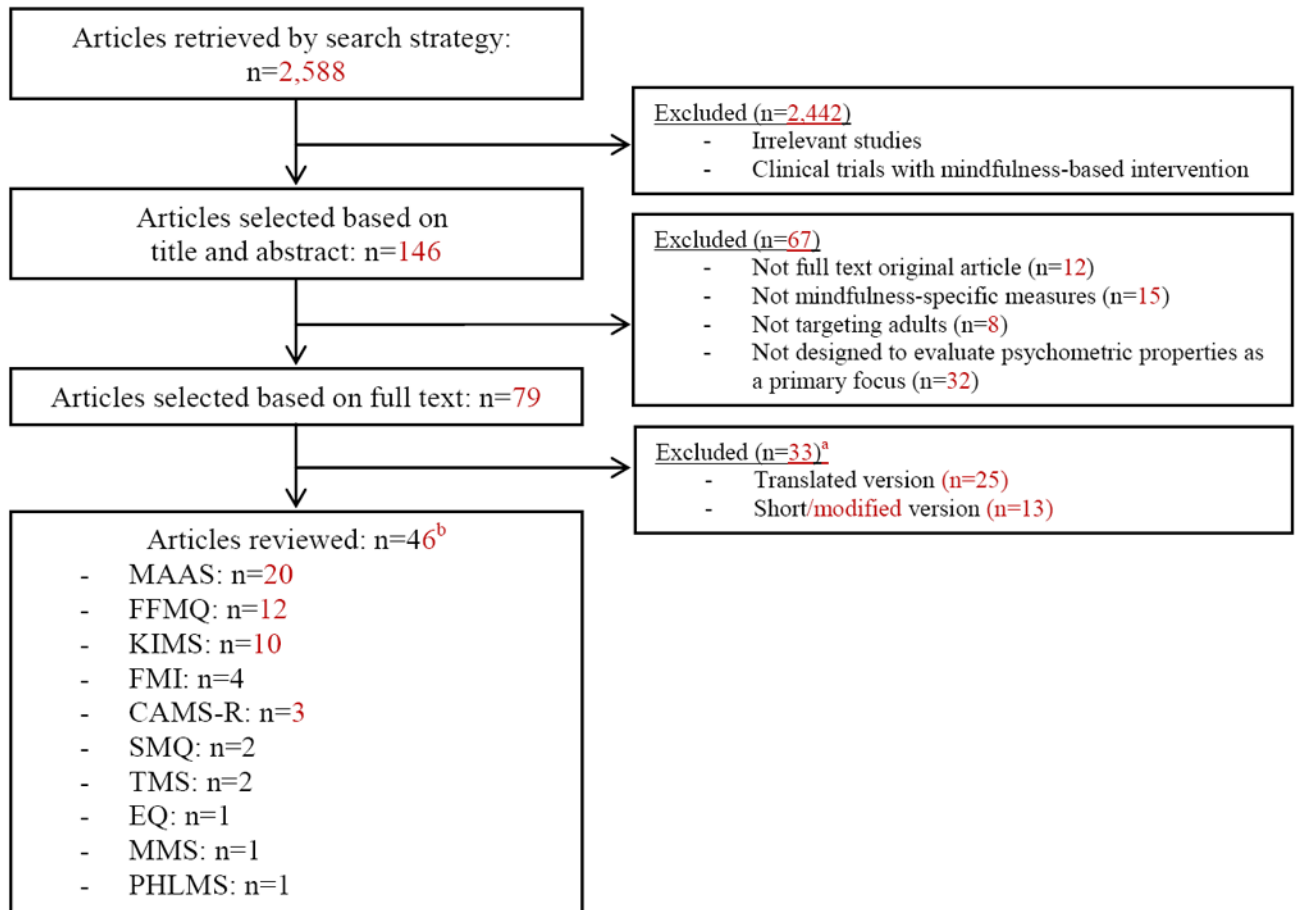
25. Grossman P. Defining mindfulness by how poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: comment on Brown et al. (2011). *Psychological Assessment*. 2011; 23(4):1034–1040. [PubMed: 22122674]
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*. 2010; 19(4):539–549. [PubMed: 20169472]
27. Framson C, Kristal AR, Schenk JM, Littman AJ, Zeliadt S, Benitez D. Development and validation of the Mindful Eating Questionnaire. *Journal of the American Dietetic Association*. 2009; 109(8): 1439–1444. [PubMed: 19631053]
28. Tharaldsen KB, Bru E. Validation of the Mindful Coping Scale. *Emotional and Behavioural Difficulties*. 2011; 16(1):87–103.
29. Reavley N, Pallant JF. Development of a scale to assess the meditation experience. *Personality and Individual Differences*. 2009; 47(6):547–552.
30. Solloway SG, Fisher WP. Mindfulness practice: A Rasch variable construct innovation. *Journal of Applied Measurement*. 2007; 8(4):359–372. [PubMed: 18250523]
31. Neff KD. The development and validation of a scale to measure self-compassion. *Self & Identity*. 2003; 2223(3)
32. Chawla N, Collin S, Bowen S, Hsu S, Grow J, Douglass A, et al. The Mindfulness-based Relapse Prevention Adherence and Competence Scale: Development, interrater reliability, and validity. *Psychotherapy Research : Journal of the Society for Psychotherapy Research*. 2010; 20(4):388–397. [PubMed: 20204916]
33. De Vet, HCW.; Terwee, CB.; Mokkink, LB.; Knol, DL. *Measurement in Medicine*. New York: Cambridge University Press; 2011.
34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010; 63(7):737–745. [PubMed: 20494804]
35. Terwee CB, Roorda LD, Knol DL, de Boer MR, de Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. *Journal of Clinical Epidemiology*. 2009; 62(10):1062–1067. [PubMed: 19230609]
36. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*. 2012; 21(4):651–657. [PubMed: 21732199]
37. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 2007; 60(1):34–42. [PubMed: 17161752]
38. van Tulder M, Furlan A, Bombardier C, Bouter L. Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine*. 2003; 28(12):1290–1299. [PubMed: 12811274]
39. Carlson LE, Brown KW. Validation of the Mindful Attention Awareness Scale in a cancer population. *Journal of Psychosomatic Research*. 2005; 58(1):29–33. [PubMed: 15771867]
40. MacKillop J, Anderson EJ. Further psychometric validation of the Mindful Attention Awareness Scale (MAAS). *Journal of Psychopathology & Behavioral Assessment*. 2007; 29(4):289–293.
41. Christopher MS, Charoensuk S, Gilbert BD, Neary TJ, Pearce KL. Mindfulness in Thailand and the United States: A case of apples versus oranges? *Journal of Clinical Psychology*. 2009; 65(6): 590–612. [PubMed: 19358288]
42. Van Dam NT, Earleywine M, Borders A. Measuring mindfulness? An item response theory analysis of the Mindful Attention Awareness Scale. *Personality & Individual Differences*. 2010; 49(7):805–810.
43. Cordon SL, Finney SJ. Measurement invariance of the Mindful Attention Awareness Scale across adult attachment style. *Measurement & Evaluation in Counseling & Development*. 2008; 40(4): 228–245.

44. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. *Assessment*. 2006; 13(1):27–45. [PubMed: 16443717]
45. Christopher MS, Gilbert BD. Incremental validity of components of mindfulness in the prediction of satisfaction with life and depression. *Current Psychology*. 2010; 29(1):10–23.
46. Ghorbani N, Watson P, Weathington B. Mindfulness in Iran and the United States: Cross-cultural structural complexity and parallel relationships with psychological adjustment. *Current Psychology*. 2009; 28(4):211–224.
47. Frewen P, Evans E, Maraj N, Dozois D, Partridge K. Letting go: Mindfulness and negative automatic thinking. *Cognitive Therapy & Research*. 2008; 32(6):758–774.
48. Waters AJ, Reitzel LR, Cinciripini P, Li Y, Marcus MT, Vidrine JI, et al. Associations between mindfulness and implicit cognition and self-reported affect. *Substance Abuse*. 2009; 30(4):328–337. [PubMed: 19904668]
49. Schmertz SK, Anderson PL, Robins DL. The relation between self-report mindfulness and performance on tasks of sustained attention. *Journal of Psychopathology and Behavioral Assessment*. 2009; 31(1):60–66.
50. Baer RA, Smith GT, Allen KB. Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment*. 2004; 11(3):191–206. [PubMed: 15358875]
51. Vujanovic AA, Youngwirth NE, Johnson KA, Zvolensky MJ. Mindfulness-based acceptance and posttraumatic stress symptoms among trauma-exposed adults without axis I psychopathology. *Journal of Anxiety Disorders*. 2009; 23(2):297–303. [PubMed: 18834701]
52. Vujanovic AA, Bonn-Miller MO, Bernstein A, McKee LG, Zvolensky MJ. Incremental validity of mindfulness skills in relation to emotional dysregulation among a young adult community sample. *Cognitive Behaviour Therapy*. 2010; 39(3):203–213. [PubMed: 20182933]
53. Buchheld N, Grossman P, Walach H. Measuring mindfulness in insight meditation (Vipassana) and meditation-based psychotherapy: The development of the Freiburg Mindfulness Inventory (FMI). *Journal of Meditation and Meditation Research*. 2001; 1(1):11–34.
54. Walach H, Buchheld N, Buttenmüller V, Kleinknecht N, Schmidt S. Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). *Personality & Individual Differences*. 2006; 40(8):1543–1555.
55. Leigh J, Bowen S, Marlatt GA. Spirituality, mindfulness and substance abuse. *Addictive Behaviors*. 2005; 30(7):1335–1341. [PubMed: 16022930]
56. Kabat-Zinn J. Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science & Practice*. 2003; 10(2):144–156.
57. Feldman G, Hayes A, Kumar S, Greeson J, Laurenceau J. Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of Psychopathology & Behavioral Assessment*. 2007; 29(3):177–190.
58. Chadwick P, Hember M, Symes J, Peters E, Kuipers E, Dagnan D. Responding mindfully to unpleasant thoughts and images: Reliability and validity of the Southampton Mindfulness questionnaire (SMQ). *The British Journal of Clinical Psychology / the British Psychological Society*. 2008; 47(Pt 4):451–455. [PubMed: 18573227]
59. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment*. 2008; 15(3):329–342. [PubMed: 18310597]
60. Christopher MS, Neuser NJ, Michael PG, Baitmangalkar A. Exploring the psychometric properties of the Five Facet Mindfulness Questionnaire. *Mindfulness*. 2012; 3(2):124–131.
61. Fernandez AC, Wood MD, Stein LA, Rossi JS. Measuring mindfulness and examining its relationship with alcohol use and negative consequences. *Psychology of Addictive Behaviors*. 2010; 24(4):608–616. [PubMed: 21198223]
62. Hollis-Walker L, Colosimo K. Mindfulness, self-compassion, and happiness in non-meditators: A theoretical and empirical examination. *Personality & Individual Differences*. 2011; 50(2):222–227.
63. Barnes SM, Lynn SJ. Mindfulness skills and depressive symptoms: A longitudinal study. *Imagination, Cognition and Personality*. 2010; 30(1):77–91.
64. Cash M, Whittingham K. What facets of mindfulness contribute to psychological well-being and depressive, anxious, and stress-related symptomatology? *Mindfulness*. 2010; 1(3):177–182.

65. Lavender JM, Gratz KL, Tull MT. Exploring the relationship between facets of mindfulness and eating pathology in women. *Cognitive Behaviour Therapy*. 2011; 40(3):174–182. [PubMed: 21660777]
66. Fisak B, von Lehe AC. The relation between the Five Facets of Mindfulness and Worry in a Non-clinical Sample. *Mindfulness*. 2012; 3(1):15–21.
67. Van Dam NT, Earleywine M, Danoff-Burg S. Differential item function across meditators and non-meditators on the Five Facet Mindfulness Questionnaire. *Personality & Individual Differences*. 2009; 47(5):516–521.
68. Baer RA, Samuel DB, Lykins EL. Differential item functioning on the Five Facet Mindfulness Questionnaire is minimal in demographically matched meditators and nonmeditators. *Assessment*. 2011; 18(1):3–10. [PubMed: 21193491]
69. Lau MA, Bishop SR, Segal ZV, Buis T, Anderson ND, Carlson L, et al. The Toronto Mindfulness Scale: Development and validation. *Journal of Clinical Psychology*. 2006; 62(12):1445–1467. [PubMed: 17019673]
70. Davis KM, Lau MA, Cairns DR. Development and preliminary validation of a trait version of the Toronto Mindfulness Scale. *Journal of Cognitive Psychotherapy*. 2009; 23(3):185–197.
71. Fresco DM, Moore MT, van Dulmen MH, Segal ZV, Ma SH, Teasdale JD, et al. Initial psychometric properties of the experiences questionnaire: Validation of a self-report measure of decentering. *Behavior Therapy*. 2007; 38(3):234–246. [PubMed: 17697849]
72. Haigh EA, Moore MT, Kashdan TB, Fresco DM. Examination of the factor structure and concurrent validity of the Langer Mindfulness/Mindlessness Scale. *Assessment*. 2011; 18(1):11–26. [PubMed: 20980699]
73. Cardaciotto L, Herbert JD, Forman EM, Moitra E, Farrow V. The assessment of present-moment awareness and acceptance: The Philadelphia Mindfulness Scale. *Assessment*. 2008; 15(2):204–223. [PubMed: 18187399]
74. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine* (1982). 1999; 48(11):1531–1548. [PubMed: 10400255]
75. Eton DT, Bauer BA, Sood A, Yost KJ, Sloan JA. Patient-reported outcomes in studies of complementary and alternative medicine: Problems, solutions, and future directions. *Explore (NY)*. 2011; 7(5):314–319. [PubMed: 21907154]
76. US Department of Health and Human Services Food and Drug Administration. US Department of Health and Human Services Food and Drug Administration; 2009. Guidance for industry: Patient-reported outcome measure: Use in medical product development to support labeling claims. <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> [11 November 2012]
77. Sauer S, Walach H, Schmidt S, Hinterberger T, Lynch S, Büssing A, et al. Assessment of mindfulness: Review on state of the art. *Mindfulness*. 2012:1–15.
78. Grossman P, Van Dam NT. Mindfulness, by any other name...: trials and tribulations of sati in western psychology and science. *Contemporary Buddhism*. 2011; 12(1):219–239.
79. Schlomer GL, Bauman S, Card NA. Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*. 2010; 57(1):1–10. [PubMed: 21133556]
80. Bohlmeijer E, ten Klooster PM, Fledderus M, Veehof M, Baer R. Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. *Assessment*. 2011; 18(3):308–320. [PubMed: 21586480]
81. Skinner TC, Robertson T, Allison GT, Dunlop S, Bucks RS. Experiential avoidance, mindfulness and depression in spinal cord injuries: A preliminary study. *Australian Journal of Rehabilitation Counselling*. 2010; 16(1):27–35.
82. Hofling V, Strohle G, Michalak J, Heidenreich T. A short version of the Kentucky Inventory of Mindfulness Skills. *Journal of Clinical Psychology*. 2011; 67(6):639–645. [PubMed: 21294117]
83. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107(2): 238–246. [PubMed: 2320703]
84. Brown, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing Structural Equation Models*. California: Sage; 1993. p. 136-162.



85. Kline, RB. Principles and practice of structural equation modeling. New York: The Guilford Press; 2004.
86. Argus G, Thompson M. Perceived social problem solving, perfectionism, and mindful awareness in clinical depression: An exploratory study. *Cognitive Therapy and Research*. 2008; 32(6):745–757.
87. Baum C, Kuyken W, Bonus M, Heidenreich T, Michalak J, Steil R. The psychometric properties of the Kentucky Inventory of Mindfulness Skills in clinical populations. *Assessment*. 2010; 17(2): 220–229. [PubMed: 20040728]
88. Bernstein A, Tanay G, Vujanovic AA. Concurrent relations between mindful attention and awareness and psychopathology among trauma-exposed adults: Preliminary evidence of transdiagnostic resilience. *Journal of Cognitive Psychotherapy*. 2011; 25(2):99–113.
89. Emanuel AS, Updegraff JA, Kalmbach DA, Ciesla JA. The role of mindfulness facets in affective forecasting. *Personality & Individual Differences*. 2010; 49(7):815–818.
90. Herndon F. Testing mindfulness with perceptual and cognitive factors: External vs. internal encoding, and the cognitive failures questionnaire. *Personality & Individual Differences*. 2008; 44(1):32–41.
91. McCracken LM, Gauntlett-Gilbert J, Vowles KE. The role of mindfulness in a contextual cognitive-behavioral analysis of chronic pain-related suffering and disability. *Pain*. 2007; 131(1-2):63–69. [PubMed: 17257755]
92. McCracken LM, Thompson M. Components of mindfulness in patients with chronic pain. *Journal of Psychopathology & Behavioral Assessment*. 2009; 31(2):75–82.
93. McKee L, Zvolensky MJ, Solomon SE, Bernstein A, Leen-Feldner E. Emotional-vulnerability and mindfulness: A preliminary test of associations among negative affectivity, anxiety sensitivity, and mindfulness skills. *Cognitive Behaviour Therapy*. 2007; 36(2):91–101. [PubMed: 17530495]
94. Roemer L, Lee JK, Salters-Pedneault K, Erisman SM, Orsillo SM, Mennin DS. Mindfulness and emotion regulation difficulties in generalized anxiety disorder: Preliminary evidence for independent and overlapping contributions. *Behavior Therapy*. 2009; 40(2):142–154. [PubMed: 19433145]
95. Vujanovic AA, Zvolensky MJ, Bernstein A, Feldner MT, McLeish AC. A test of the interactive effects of anxiety sensitivity and mindfulness in the prediction of anxious arousal, agoraphobic cognitions, and body vigilance. *Behaviour Research and Therapy*. 2007; 45(6):1393–1400. [PubMed: 16949030]
96. Zvolensky MJ, Solomon SE, McLeish AC, Cassidy D, Bernstein A, Bowman CJ, et al. Incremental validity of mindfulness-based attention in relation to the concurrent prediction of anxiety and depressive symptomatology and perceptions of health. *Cognitive Behaviour Therapy*. 2006; 35(3): 148–158. [PubMed: 16952898]



**Figure 1.**

Flowchart of the search strategy and selection of articles

<sup>a</sup> Several papers included both translated and created a short/or modified version.

<sup>b</sup> Several papers reviewed multiple scales and/or contained multiple studies.

Table 1

Quality criteria for measurement properties adapted from Terwee et al. [37]

Property	Rating	Quality Criteria <sup>a</sup>
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	-	MIC ≤ SDC OR MIC equals or inside LOA
Reliability	+	ICC/weighted Kappa 0.70 OR Pearson's <i>r</i> 0.80
	?	Neither ICC/weighted Kappa, nor Pearson's <i>r</i> determined
	-	ICC/weighted Kappa < 0.70 OR Pearson's <i>r</i> < 0.80
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement OR <i>no assessment of completeness or comprehensiveness</i> <sup>a</sup>
	-	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity		
Structural validity	+	Factors should explain at least 50% of the variance OR <i>Good or adequate fit by goodness-of-fit criteria for a CFA or EFA</i> <sup>a, b</sup>
	?	Explained variance not mentioned OR <i>equivocal fit by goodness-of-fit criteria for a CFA or EFA</i> <sup>a, b</sup>
	-	Factors explain < 50% of the variance OR <i>Poor fit by goodness-of-fit criteria for a CFA or EFA</i> <sup>a, b</sup>
Hypothesis testing	+	Correlation with an instrument measuring the same construct 0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs OR <i>No evidence of DIF</i> <sup>a</sup>
	?	Solely correlations determined with unrelated constructs OR <i>50% but &lt; 75% of the results are in accordance with the hypotheses</i> <sup>a</sup> OR <i>possible DIF</i> <sup>a</sup>
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 50% of the results are in accordance with the hypotheses <sup>a</sup> OR correlation with related constructs is lower than with unrelated constructs OR <i>notable evidence of DIF</i> <sup>a</sup>
Responsiveness		
Responsiveness	+	Correlation of changes with an instrument measuring change in the same construct 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC 0.70 AND correlation of changes with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation of changes with an instrument measuring change in the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation of changes with related constructs is lower than with unrelated constructs

MIC: Minimal important change, SDC: Smallest detectable change, LOA: Limits of agreement, ICC: Intraclass correlation coefficient, DIF: Differential item functioning, AUC: Area under the curve

+ positive rating, ? indeterminate rating, - negative rating

<sup>a</sup>Quality criteria are from [37], with modifications by the authors indicated by italics.

<sup>b</sup> Good or Adequate fit: comparative fit index (CFI) = 0.90, root mean square error of approximation (RMSEA) = 0.08, standardized root mean square residual (SRMR) < 0.10 [83-85]; Inadequate fit: CFI = 0.85, RMSEA = 0.10, SRMR = 0.10; Indeterminate fit: the values of fit indexes ranged in between the adequate criteria and inadequate criteria.

**Table 2**

Levels of evidence for the overall quality of the measurement properties

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of <i>good</i> methodological quality OR in one study of <i>excellent</i> methodological quality
Moderate	++ or --	Consistent findings in multiple studies of <i>fair</i> methodological quality OR in one study of <i>good</i> methodological quality
Limited	+ or -	One study of <i>fair</i> methodological quality
Conflicting	±	Conflicting findings from studies of comparable quality
Indeterminate	?	Findings from <i>excellent</i> , <i>good</i> or <i>fair</i> studies were not definitively positive or negative
None	na	Findings from <i>excellent</i> , <i>good</i> or <i>fair</i> were not available

Table adapted from Van Tulder et al. [38]: + positive result; - negative result; ± both positive and negative findings have been reported by studies of adequate quality; ? findings from studies of adequate quality were not definitively positive or negative; na findings from studies of adequate quality were not available

Table 3

## Characteristics of the included studies

Study	Population	Sample size	Age, mean (SD)	Female (%)	Country
Argus and Thomson (2008) [86]	Patients with depression	141	43 (14)	68	Australia
Baer et al. (2004) [50]	Substudy 1: Psychologists/Doctoral students in clinical psychology Substudy 2 and substudy 6: Undergraduates (sample 1)/Undergraduates (sample 2)/Adults with borderline personality disorder (sample 3) Substudy 3: Subset of sample 2 Substudy 4: Subset of sample 1 Substudy 5: Undergraduates	5/6 205/215/26 49 130 115	nr mostly 18-22 (nr)/ nr/36 (nr) nr 20 (nr) nr	100/50 about 60/nr/96 nr 56 nr	US US US US US
Baer et al. (2006) [44]	Substudy 1 and substudy 2: Undergraduates Substudy 3: Undergraduates Substudy 4: Samples from substudy 1, substudy 2, and substudy 3	613 268 881	21 (nr) 19 (nr) 20 (nr)	70 77 72	US US US
Baer et al. (2008) [59]	Undergraduates/Community participants/Nonmeditators/Meditators	259/293/252/213	19 (3)/50 (7)/44 (12)/49 (13)	78/60/58/68	US/UK/US/US
Baer et al. (2011) [68]	Meditators Nonmeditators Undergraduates	115 115 145	46 (12) 44 (12) 19 (2)	73 63 69	US US US
Baum et al. (2010) [87]	Patients with recurrent depression	100	48 (11)	77	England
Bernstein et al. (2011) [88]	Adults with traumatic events	76	30 (13)	46	US
Brown and Ryan (2003) [15]	Study: Undergraduates/Undergraduates/General adults/Undergraduates Substudy 1: Undergraduates, general adults Substudy 2: Community meditators and nonmeditators Substudy 3: Undergraduates Substudy 4: Community participants/Undergraduates Substudy 5: Patients with breast or prostate cancer	313/327/239/60 74 to 1,046 100 90 74/92 41 115	20 (nr)/20 (nr)/43 (nr)/19 (nr) 19 to 23 (nr) 41 (nr) 20 (nr) 38 (nr)/20 (nr) 55 (10) 43 (nr)	66/64/66/57 55 to 66 29 66 55/74 78 69	US US US US US US German
Buchheld et al. (2001) [53]	Retreat participants	6	nr	33	US
Cardaciottio et al. (2008) [73]	Substudy 1: Mindfulness experts Substudy 2: Undergraduates Substudy 3: Undergraduates Substudy 4: Outpatients in psychiatry	204 559 52	22 (4) 20 (4) 41 (12)	52 51 56	US US US US

Study	Population	Sample size	Age, mean (SD)	Female (%)	Country
	Substudy 5: Inpatients with eating disorders	30	30 (11)	90	US
	Substudy 6: Graduate students seeking psychotherapy	78	26 (8)	89	US
Carlson and Brown (2005) [39]	Outpatients with cancer	122	50 (13)	67	Canada
	Community participants	122	48 (16)	67	Canada
Cash and Whittingham (2010) [64]	Community participants	106	36 (16)	59	Australia
Chadwick et al. (2008) [58]	Community participants: Nonmeditators	51	47 (nr)	75	England
	Community participants: Meditators	83	47 (nr)	60	England
	Patients with psychosis	122	31 (nr)	36	England
Christopher et al. (2009) <sup>a</sup> [41]	Undergraduates	365	22 (6)	71	US
Christopher and Gilbert (2010) [45]	Undergraduates	365	22 (6)	71	US
Christopher et al. (2012) [60]	Adults online (meditators and nonmeditators)	349	32 (12)	75	US
Cordon and Finney (2008) [43]	Securely attached undergraduates	228	19 (1)	77	US
	Insecurely attached undergraduates	267	19 (1)	77	US
Davis et al. (2009) [70]	Community participants	369	40 (14)	65	Australia
	Undergraduates	92	22 (7)	80	Australia
Emanuel et al. (2010) [89]	Undergraduates: online responders	109	23 (nr)	78	US
	Undergraduates: paper responders	111	nr	nr	US
Feldman et al. (2007) [57]	Substudy 1: Undergraduates/Undergraduates	250/298	19 (3)/19 (2)	64/61	US
	Substudy 2: Undergraduates	212	19 (2)	60	US
Fernandez et al. (2010) [61]	Undergraduates	316	22 (0.4)	92	US
Fisak and von Lehe (2012) [66]	Undergraduates	400	22 (5)	69	US
Fresco et al. (2007) [71]	Substudy 1: Undergraduates/Undergraduates	1,150/519	19 (4)/19 (2)	67/65	US/US
	Substudy 2: Undergraduates	61	20 (3)	56	US
	Substudy 3: Patients with depression vs. Healthy control	220 vs. 50	44 (10) vs. 45 (9)	75 vs. 74	England, Wales, Canada vs. England
Frewen et al. (2008) [47]	Substudy 1: Undergraduates	64	nr	73	Canada
	Substudy 2: Undergraduates	43	nr	70	Canada
	Undergraduates: three samples	256/298/346	20 (5)/19 (3)/20 (3)	37/68/54	US
Ghorbani et al. (2009) <sup>a</sup> [46]	Undergraduates (sample 1)/Undergraduates (sample 2)/ Undergraduates (subset of sample 1 + subset of sample 2)	582/457/451	19 (4)/23 (10)/22 (nr)	69/60/69	US

Study	Population	Sample size	Age, mean (SD)	Female (%)	Country
Hemdon (2008) [90]	Undergraduates	142	nr	nr	US
Hollis-Walker and Colosimo (2011) [62]	Undergraduates and demographically similar community participants	123	21 (nr)	78	Canada
Lau et al. (2006) [69]	Substudy 1: General adults (Meditators + Nonmeditators)	390 (232+158)	41 (13)	55	Canada
Lavender et al. (2011) [65]	Substudy 2: Patients with psychiatric or medical conditions	123	47 (13)	68	Canada
Leigh et al. (2005) [55]	Undergraduate women	276	20 (3)	100	US
MacKillop and Anderson (2007) [40]	Undergraduates	196	nr	63	US
McCracken et al. (2007) [91]	Undergraduates	711	Mostly 18-19 (nr)	53	US
McCracken et al. (2007) [91]	Patients with pain	105	47 (13)	60	UK
McCracken and Thompson (2009) [92]	Patients with pain	150 <sup>b</sup>	47 (13)	64	UK
McKee et al. (2007) [93]	Community participants	154	22 (8)	57	US
Roemer et al. (2009) [94]	Substudy 1: University commuters	411	23 (nr)	64	US
Schmertz et al. (2009) [49]	Substudy 2: Clinical sample diagnosed with generalized anxiety disorder vs. control	16 vs. 16	33 (12) vs. 31 (9)	69 vs. 69	US
Van Dam et al. (2009) [67]	Undergraduates	50	20 (3)	82	US
Van Dam et al. (2010) [42]	Undergraduates: Nonmeditators/Meditators	263/58	19 (1)/48 (14)	48/64	US
Vujanovic et al. (2007) [95]	Undergraduates	414	20 (3)	67	US
Vujanovic et al. (2009) [51]	Community participants	248	22 (8)	55	US
Vujanovic et al. (2010) [52]	Adults with traumatic life events	239	23 (10)	54	US
Walach et al. (2006) [54]	Non-clinical community young adults	193	24 (10)	55	US
Waters et al. (2009) [48]	Meditators/General adults/Clinical sample with psychiatric conditions	85/85/117	44 (9)/34 (12)/nr	66/55/nr	German
Zvolensky et al. (2006) [96]	Adult smokers	158	44 (12)	45	US
	Community participants	170	22 (8)	56	US

nr: Not reported

<sup>a</sup> Studies were limited to English speaking students.

<sup>b</sup> Includes the sample from McCracken et al. 2007; the 2009 findings addressed in this review were limited to properties not assessed in the 2007 report.



## Characteristics of the included instruments

Table 4

Instrument	Construct assessed <sup>a</sup>	Rec all period	Dimensions (number of items)	Number of subscales	Response options (range)	Ease of scoring and administration (Range of scores)	Sample items
MAAS	Dispositional Mindfulness	None	One dimension Total (15)	None	6-point scale (1 = almost always to 6 = almost never)	Easy <sup>b</sup> (1-6)	"I find myself doing things without paying attention," and "I do jobs or tasks automatically, without being aware of what I'm doing."
KIMS	Mindfulness Skills	None	Observe (12) Describe (8) Act with awareness (10) Accept without judgment (9) Total (39)	4	5-point scale (1 = never or very rarely true to 6 = almost always or always true)	Easy (Observe: 12-72, Describe: 8-48, Act with awareness: 10-60, Accept without judgment: 9-54)	"I notice the smells and aromas of things" (Observe); "I'm good at finding the words to describe my feelings" (Describe); "When I do things, my mind wanders off and I'm easily distracted" (Act with Awareness); and "I tell myself that I shouldn't be feeling the way I'm feeling" (Accept without Judgment).
FMI	Mindfulness	Time frame to be set by user	Present-moment disidentifying attention (12) Nonjudgmental, nonevaluative attitude toward self and others (7) Openness to negative mind states (7) Process-oriented, insightful understanding (4) Total (30)	None	4-point scale (1 = almost never to 4 = almost always)	Easy (30-120)	"I am open to the experience of the present moment," and "I perceive my feelings and emotions without having to react to them."
CAMS-R	Mindfulness	None	Attention (3) Present-focus (3)	None	4-point scale (1 = rarely/not at all to 4 = almost always)	Easy (12-48)	"It is easy for me to concentrate on what I am doing,"

Instrument	Construct assessed <sup>a</sup>	Rec all period	Dimensions (number of items)	Number of subscales	Response options (range)	Ease of scoring and administration (Range of scores)	Sample items
SMQ	Mindfulness	None	Awareness (3) Acceptance/non-judgment (3) Total (12) One dimension Total (16)	None	7-point scale (0 = strongly disagree to 6 = strongly agree)	Easy (0-96)	and "I am able to focus on the present moment,"  All items start with "Usually when I experience distressing thoughts and images..." and items include, "I am able just to notice them without reacting," and "I am able to accept the experience."
FFMQ	Mindfulness	None	Observing (8) Describing (8) Acting with awareness (8) Nonjudging of experience (8) Nonreactivity to experience (7) Total (39)	5	5-point scale (1 = never or very rarely true to 5 = very often or always true)	Easy (Observing: 8-40, Describing: 8-40, Acting with awareness: 8-40, Nonjudging of experience: 8-40, Nonreactivity to experience: 7-35)	"I pay attention to sensations such as the wind in my hair or sun on my face" (Observing); "I have trouble thinking of the right words to express how I feel about things," (Describing); "I rush through activities without being really attentive to them" (Acting with awareness); "I make judgments about whether my thoughts are good or bad" (Nonjudging of experience); and "I watch my feelings without getting lost in them" (Nonreactivity to experience).
TMS	Mindfulness	Respondents are instructed	Curiosity (6)	2	5-point scale (0 = not at all to 4 = very much)	Moderate <sup>c</sup> (Curiosity: 0-24, Decentering: 0-28)	"I was curious about my

Instrument	Construct assessed <sup>a</sup>	Rec all period	Dimensions (number of items)	Number of subscales	Response options (range)	Ease of scoring and administration (Range of scores)	Sample items
		to rate a 15 minute meditation experience	Decentering (7)				reactions to things," (Curiosity) and "I was aware of my thoughts and feelings without over-identifying with them" (Decentering).
EQ	Decentering	None	One dimension Total (11)	None	5-point scale (1 = never to 5 = all the time)	Easy (11-55)	"I am better able to accept myself as I am," and "I can observe unpleasant feelings without being drawn into them."
MMS	Mindfulness	None	Novelty seeking (6) Novelty producing (6) Engagement (5) Flexibility (4) Total (21)	4	7-point scale (1 = strongly disagree to 7 = strongly agree)	Easy (Novelty seeking: 6-42, Novelty producing: 6-42, Engagement: 5-35, Flexibility: 4-28)	"I like to investigate things" (Novelty seeking); "I try to think of new ways of doing things" (Novelty producing); "I get involved in almost everything I do" (Engagement); and "I stay with the old tried and true ways of doing things" (Flexibility).
PHLMS	Mindfulness	One week	Awareness (10) Acceptance (10)	2	5-point scale (1 = never to 5 = very often)	Awareness: easy (10-50) Acceptance: easy <sup>d</sup> (10-50)	"When I am startled, I notice what is going on inside my body," (Awareness) and "There are things I try not to think about" (Acceptance).

<sup>a</sup> Higher scores represent higher levels of construct.

<sup>b</sup> Average of total summed items

<sup>c</sup> It requires a delivery of 15 minute meditation period prior to administration.

$d$  Sum of the reversed scores

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**Table 5**

Methodological quality of each study per measurement property and instrument

	Internal consistency	Reliability	Content validity	Structural validity	Hypotheses testing	Responsiveness
<b>MAAS</b>						
Brown and Ryan (2003)	Good <sup>c</sup>	Good <sup>c</sup>	Fair <sup>d</sup>	Good <sup>b</sup>	Fair <sup>c</sup>	
Brown and Ryan – substudy 1 (2003)					Good <sup>c</sup>	
Brown and Ryan – substudy 2 (2003)					Good <sup>c</sup>	
Brown and Ryan – substudy 3 (2003)					Good <sup>c</sup>	
Brown and Ryan – substudy 4 (2003)					Good <sup>c</sup>	
Brown and Ryan – substudy 5 (2003)	Fair <sup>c</sup>					Fair <sup>c</sup>
Carlson and Brown (2005)	Good <sup>c</sup>			Good <sup>c</sup>	Good <sup>c</sup>	
Baer et al. (2006)	Good <sup>c</sup>				Good <sup>c</sup>	
Zvolensky et al. (2006)					Fair <sup>c</sup>	
MacKillip and Anderson (2007)	Excellent <sup>c</sup>			Excellent <sup>c</sup>	Good <sup>c</sup>	
McCracken et al. (2007)	Good <sup>c</sup>				Good <sup>c</sup>	
Vujanovic et al. (2007)					Fair <sup>c</sup>	
Argus and Thomson (2008)	Good <sup>c</sup>				Fair <sup>c</sup>	
Cordon and Finney (2008)	Good <sup>c</sup>			Good <sup>c</sup>	Fair <sup>c</sup>	
Frewen et al – substudy 1 (2008)	Good <sup>c</sup>	Poor <sup>c</sup>			Good <sup>c</sup>	
Frewen et al – substudy 2 (2008)	Fair <sup>c</sup>	Poor <sup>c</sup>			Fair <sup>c</sup>	
Hemdon (2008)	Fair <sup>c</sup>				Fair <sup>c</sup>	
Christopher et al. (2009)				Fair <sup>c</sup>		
Ghorbani et al. (2009)	Good <sup>c</sup>			Good <sup>c</sup>	Good <sup>c</sup>	
McCracken and Thompson (2009)				Good <sup>c</sup>		
Roemer et al. – substudy 1 (2009)	Good <sup>c</sup>				Good <sup>c</sup>	
Roemer et al. – substudy 2 (2009)	Fair <sup>c</sup>				Fair <sup>c</sup>	
Schmertz et al. (2009)	Good <sup>c</sup>				Fair <sup>c</sup>	

	Internal consistency	Reliability	Content validity	Structural validity	Hypotheses testing	Responsiveness
Waters et al. (2009)					Good <sup>c</sup>	
Christopher and Gilbert (2010)	Good <sup>c</sup>				Good <sup>c</sup>	
Van Dam et al. (2010)	Excellent <sup>c</sup>			Fair <sup>c</sup>		
Bernstein et al. (2011)					Good <sup>c</sup>	
<b>KIMS</b>						
Baer et al. – substudy 1 (2004)			Fair <sup>d</sup>			
Baer et al. – substudy 2 (2004)	Good <sup>e</sup>			Good <sup>e</sup>		
Baer et al. – substudy 3 (2004)		Fair <sup>e</sup>				
Baer et al. – substudy 4 (2004)					Fair <sup>e</sup>	
Baer et al. – substudy 5 (2004)					Good <sup>e</sup>	
Baer et al. – substudy 6 (2004)					Fair <sup>e</sup>	
Baer et al. (2006)	Good <sup>e</sup>				Good <sup>e</sup>	
McKee et al. (2007)					Good <sup>e</sup>	
Frewen et al – substudy 1 (2008)	Good <sup>e</sup>	Poor <sup>e</sup>			Fair <sup>e</sup>	
Frewen et al – substudy 2 (2008)	Fair <sup>e</sup>	Poor <sup>e</sup>			Fair <sup>e</sup>	
Christopher et al. (2009)				Fair <sup>e</sup>		
Vujanovic et al. (2009)					Good <sup>e</sup>	
Waters et al. (2009)					Good <sup>e</sup>	
Baum et al. (2010)	Good <sup>e</sup>					
Christopher and Gilbert (2010)	Good <sup>e</sup>				Good <sup>e</sup>	
Vujanovic et al. (2010)					Good <sup>e</sup>	
<b>FMI</b>						
Buchheld et al. (2001)	Good			Fair	Fair	
Leigh et al. (2005)	Good			Good		
Baer et al. (2006)	Good				Good	
Walach et al. (2006)	Excellent			Good	Good	
<b>CAMS-R</b>						
Baer et al. (2006)	Good				Good	

	Internal consistency	Reliability	Content validity	Structural validity	Hypotheses testing	Responsiveness
Feldman et al. – substudy 1 (2007)	Good			Good		
Feldman et al. – substudy 2 (2007)	Good				Good	
Schmertz et al. (2009)	Fair				Fair	
<b>SMQ</b>						
Baer et al. (2006)	Good				Good	
Chadwick et al. (2008)	Good			Good	Fair	
<b>FFMQ</b>						
Baer et al. (2006)	Good			Good	Good	
Baer et al. (2008)	Good			Good	Good	
Van Dam et al. (2009)	Poor				Fair	
Barns and Lynn (2010-2011)	Good				Fair	
Cash and Whittingham (2010)					Good	
Emanuel et al. (2010)	Good				Fair	
Fernandez et al. (2010)	Good			Good	Fair	
Baer et al. (2011)					Good	
Hollis-Walker and Colosimo (2011)					Good	
Lavender et al. (2011)	Good				Good	
Christopher et al. (2012)	Good			Good	Fair	
Fisak and von Lehe (2012)	Good				Good	
<b>TMS</b>						
Lau et al. (2006)			Poor			
Lau et al. – substudy 1 (2006)	Good			Good	Good	
Lau et al. – substudy 2 (2006)					Good	Good
Davis et al. (2009)	Good				Good	
<b>EQ</b>						
Fresco et al. – substudy 1 (2007)	Good			Good		
Fresco et al. – substudy 2 (2007)					Good	
Fresco et al. – substudy 3 (2007)	Good			Good	Good	
<b>MMS</b>						
Haigh et al. (2011)	Excellent			Excellent		
<b>PHLMS</b>						

	Internal consistency	Reliability	Content validity	Structural validity	Hypotheses testing	Responsiveness
Cardaciotto et al. – substudy 1 (2008)			Fair <sup>f</sup>			
Cardaciotto et al. – substudy 2 (2008)	Good <sup>g</sup>					
Cardaciotto et al. – substudy 3 (2008)	Good <sup>g</sup>		Good <sup>g</sup>		Good <sup>g</sup>	
Cardaciotto et al. – substudy 4 (2008)	Fair <sup>g</sup>				Good <sup>g</sup>	
Cardaciotto et al. – substudy 5 (2008)	Fair <sup>g</sup>				Fair <sup>g</sup>	
Cardaciotto et al. – substudy 6 (2008)	Fair <sup>g</sup>				Good <sup>g</sup>	

Note: The original COSMIN checklist was modified so that studies could receive an overall rating of *good*, if the only flaw noted was inadequate reporting of the methods for handling missing data.

<sup>a</sup> Assessed with a pool of 184 items,

<sup>b</sup> with a set of 24 items,

<sup>c</sup> and with the final 15 items

<sup>d</sup> Assessed with a pool of 77 items,

<sup>e</sup> and with the final 39 items

<sup>f</sup> Assessed with a pool of 105 items,

<sup>g</sup> and with the final 20 items



**Table 6**

Quality of measurement properties per instrument<sup>a</sup>

Instrument	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing	Responsiveness
MAAS	+++	++	?	±	+++	+
KIMS	+++	+ <sup>b</sup>	?	?	++ <sup>c</sup>	na
FMI	+++	na	na	?	±	na
CAMS-R	++	na	na	++	++	na
SMQ	+++	na	na	--	±	na
FFMQ	+++	na	na	+	+++ <sup>d</sup>	na
TMS	+++	na	?	++	++	<sup>g</sup> e
EQ	+++	na	na	+++	+++	na
MMS	+++ <sup>f</sup>	na	na	---	na	na
PHLMS	+++	na	?	++	±	na

<sup>a</sup>There was no evidence to evaluate measurement error for these instruments.

<sup>b</sup>Test-retest results were adequate for 3 of 4 subscales.

<sup>c</sup>Rating for the global score is ++, ratings for subscales range from +++ for Accept without Judgment to – for Observe.

<sup>d</sup>Ratings for subscales Observe and Describe are less often consistently positive than ratings for the other subscales.

<sup>e</sup>One of the two subscales has been shown to be responsive.

<sup>f</sup>This was evaluated with the whole MMS score only.