



Research article

Development and validation of 21-item outcome inventory (OI-21)

Nahathai Wongpakaran^a, Tinakon Wongpakaran^{a,*}, Zsuzsanna Kövi^{b,**}^a Department of Psychiatry, Faculty of Medicine, Chiang Mai University, 50200, Thailand^b Institute of Psychology, Centre of Specialist Postgraduate Programmes in Psychology, Károli Gáspár University of the Reformed Church in Hungary, Budapest, Hungary

ARTICLE INFO

Keywords:

Psychiatric symptoms
Self-report
Psychometric
Outcome
Screening
Measurement
Psychopathology

ABSTRACT

Background: Outcome measurement is important for monitoring patients' progress. The study aimed to develop an outcome inventory (OI) for clinical use in routine practice in psychiatric services and to examine the psychometric properties of the newly developed OI.**Methods:** 48 items measuring anxiety, depression, interpersonal difficulties, and somatization were collected. Factor analysis was used to reduce the number of items. The final OI consisting of 21 items was then examined for psychometric properties among 1302 participants, 880 were nonclinical and 422 clinical patients. Tests included confirmatory factor analysis, internal consistency, test-retest reliability, convergent and discriminant validity, and diagnostic ability for major depression. Responsiveness was compared between baseline and 3-month follow-up.**Results:** Confirmatory factor analysis revealed the OI-21 demonstrated the designated four components. Cronbach's alpha was good to excellent for all subjects with good test-retest reliability, concurrent validity, convergent and discriminant validity. It demonstrated area under the ROC curve of 0.89 indicating good diagnostic performance. Sensitivity to change after 3 months was observed in both types of treatment. However, interpersonal difficulties were sensitive to change in those receiving additional psychotherapy.**Conclusion:** OI-21 demonstrated its validity, reliability, and sensitivity to change. It constitutes a promising tool for outcome assessment in nonclinical populations and among psychiatric patients.

1. Introduction

Along with clinician-rated, self-report measurement provides data concerning psychological distress and psychopathology in broader and more coverage than usual. Also, it helps clinicians to gain greater awareness of patients' problems as well as to effectively monitor patients' progress without any burden on the therapist to endeavor to administer the measurement (Carlier et al., 2012). The importance of using self-report questionnaires in psychiatry practice is strikingly called forth when they are included in the DSM-5 (American Psychiatric, 1996; Skodol, 2011; Skodol and Bender, 2009; Stein et al., 2009; Trull et al., 2011), in hoping that it would demonstrate treatment or intervention more apparently when combining a more dimensional approach of self-report measurement with DSM's set of categorical diagnoses. For instance, depressive symptoms assessed by self-report questionnaires such as Depression Anxiety Stress Scales (DASS) (Lovibond and Lovibond, 1995), Outcome Questionnaire (OQ)-45 (Lambert et al., 2004), Beck Depression Inventory (BDI) (Beck et al., 1961) and Geriatric depression scale (GDS) (Yesavage et al., 1982) can be useful in

monitoring the change of clinical symptom of depression regardless of the clinical status of the patient at that time, e.g., remitted, partial response.

Many self-report measurements have been used in routine monitoring of clinical practice to provide more information for decision making for clinicians. Discrepancies between clinician rated and self-report on the symptoms is common (Wongpakaran et al., 2013). A study regarding inpatients' improvement from depression showed the importance of including the patient self-reports along with clinician rating because in the case of nonresponse and deterioration, the clinical impression of change in symptom severity is often inaccurate and does not match the patient's perspective (Kaiser et al., 2022). On the other hand, another study suggested discrepancy between self-report and clinician rating on depressive symptom are low among patients in remission. As such, it would be sufficient to use the self-report version of a questionnaire to screen, monitor, and detect remission for MDD symptoms (Lyu et al., 2019). This suggests that self-report questionnaires might be equal to or even better than clinician rating in evaluating the patient's actual feelings. This evidence highlights the merit of self-report questionnaires.

* Corresponding author.

** Corresponding author.

E-mail addresses: tinakon.w@cmu.ac.th (T. Wongpakaran), kovi.zsuzsanna@kre.hu (Z. Kövi).<https://doi.org/10.1016/j.heliyon.2022.e09682>

Received 4 January 2022; Received in revised form 4 March 2022; Accepted 1 June 2022

2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The common self-report questionnaires used in clinical or nonclinical settings include Clinical Outcomes in Routine Evaluation-Outcome Measures (CORE-OM) (Barkham et al., 2001), OQ-45 (Ellsworth et al., 2006), Symptom Checklist-90 (SCL-90) (Derogatis et al., 1973), Brief Symptom Inventory (BSI, BSI-18) (Recklitis and Rodriguez, 2007), Symptom Questionnaire (SQ-48), Mood and Anxiety Symptoms Questionnaire (MASQ) (Wardenaar et al., 2010) and General Health Questionnaire (GHQ) (Goldberg and Williams, 1988), whereas some are used for psychotherapy settings such as OQ-45 (Lo Coco et al., 2008), the Depression Anxiety Stress Scales (DASS) (Lovibond and Lovibond, 1995), OQ-45, Behaviour and Symptom Identification-24 (BASIS-24) (Cameron et al., 2007) and SCL-90-R/BSI (Tarescavage and Ben-Porath, 2014).

As aforementioned, it has become evident that self-report measures can cover a wide range of symptoms. Among various symptoms used for clinical practice, anxiety and depression are of interest to clinicians and can be observed in nearly all patients, regardless of diagnosis and severity. These two common symptoms are the fundamental indicators to be measured in any setting. In addition to anxiety and depression, somatization is considered one of the main symptoms and is usually part of anxiety and depression. It constitutes a separately diagnosed somatic symptom disorder in the DSM-5 (APA, 2013). Therefore, somatization should also be included in the symptom questionnaire.

Anxiety, depression and somatization are also among the most common psychiatric disorders (Hanel et al., 2009; Kohlmann et al., 2016; Löwe et al., 2008; Means-Christensen et al., 2008). Apart from the symptoms, other domains assessed and monitored include social role subscale (in OQ); well-being, function and risk (in CORE-OM) and impulsive/addictive behaviors (in BASIS). What problems or symptoms to be included in a measurement is contingent on an prevalence of problems or symptoms in that given area. For example, Outcome Questionnaire-45, contains items related to drug abuse because substance abuse is common and produces challenging issues in that area where the development of the questionnaire takes place (Lo Coco et al., 2008; Nebeker et al., 1995; Whipple and Lambert, 2011).

Routine clinical practice is mainly an outpatient service where psychotropic medication as well as brief counseling regarding medication are the focus. Studies have demonstrated that symptom distress changes more quickly and strongly than interpersonal problems (Liebherz and Rabung, 2014) (Berghout et al., 2012). The measurement should include symptoms that are sensitive to be changed, e.g., anxiety and depression to demonstrate the effectiveness of medication treatment. When it comes to psychotherapy, items that can demonstrate the effectiveness of psychotherapy should be included such as interpersonal difficulties. These items are difficult to change by medication treatment, and considered as a trait rather than symptoms (Quilty et al., 2013). The same is true for somatization symptoms in which psychotherapy rather than medication is shown to be effective for these symptoms (Allen et al., 2006; van Dessel et al., 2014). Therefore, to detect the effectiveness of psychotherapy, slow to change symptoms such interpersonal difficulties and somatization should be included in the questionnaire.

In terms of items representing each symptom, less concern was shown regarding anxiety and depressive symptoms. However, when it comes to somatization symptom and interpersonal problem, the list of the symptoms to be included in the questionnaire depends on the epidemiology in that targeted population. Our related study demonstrated the prevalence of some specific somatization symptoms (Wongpakaran et al., 2011). Likewise, interpersonal problems and submissive interpersonal problems were prominent among Thais compared with US subjects (Wongpakaran et al., 2012).

Another point to be considered is the items related to cultural bias. Differential item functioning can be found due to culture (Forero et al., 2014). Such evidence has been shown in many existing measurements. For example, regarding depressive symptoms, the differential item functioning due to culture was noted in PHQ-9 (Reich et al., 2018), GDS (Broekman et al., 2008; Wongpakaran et al., 2019), BDI (Azocar et al., 2001), CESD (Choi et al., 2009), 4DSQ (Terluin et al., 2016) as well as

other measurements, especially somatic symptom related to depression (Kalibatseva et al., 2014; Kirmayer, 2001; Uebelacker et al., 2009). Evidence of cultural bias was found in anxiety symptom (Giannopoulou et al., 2021; Hoge et al., 2006; Kirmayer, 2001; Parkerson et al., 2015; Wiesner et al., 2010), likewise, in somatization (Kirmayer and Young, 1998; Wiesner et al., 2010). Moreover, studies showed a problem of cross-cultural validity in some well-validated measurements. A cultural bias could influence in cross-cultural translation or adaptation processes (Kemmelmeier, 2016; Rompel et al., 2017). Some well-validated measurement may not assess problems faced in other areas or between different cultures. Therefore, to create a measurement containing specific items corresponding to the problems for the targeted population was one way to avoid item bias that usually occurred when using a measurement from different cultures. In sum, the rationale to have this new scale was to have a scale that 1) was designed for the authors' symptoms of interest, that covered four dimensions, 2) was able to identify symptoms specific to targeted population, and 2) contained the possible bias-free items, e.g., cultural.

Lastly, many self-report measurements are not free of charge, e.g., OQ-45, BDI, while many have been created for public domain. Our attempt to create this measurement is consistent with growing efforts in other scientific areas to develop tools that are free of charge (Moessner et al., 2011), e.g., PHQ-9 (depression), GAD-7 (anxiety) and PHQ-15 (somatization). Consistent with other questionnaires such as 4DSQ (Terluin et al., 2006) and SQ-48 (Carlier et al., 2012), the new tool should cover 3 to 4 main symptoms to save time for administration, while it was also determined not to be too lengthy a questionnaire.

The purposes of this study were to develop an outcome inventory (OI) for clinical use in routine practices in mental health, e.g., psychiatric services and psychotherapy settings, and to investigate its psychometric properties. In addition, OI-21 was developed as a public domain instrument, freely available to researchers and clinicians. The study sought to examine the OI for its construct validity, convergent and discriminant validity, concurrent validity and sensitivity to change. In terms of reliability, the study explored internal consistency using Cronbach's alpha and test-retest reliability.

2. Materials and methods

We designed the OI-21 in three stages. First was item development. Second, we evaluated its psychometric properties including construct validity, reliability, and test-retest reliability among nonclinical subjects, and lastly, we assessed its sensitivity to change in a clinical sample by comparing the group with treatment as usual and the group with psychotherapeutic intervention.

2.1. Participants

2.1.1. Nonclinical participants

This group of subjects comprised nonclinical participants, i.e., university students, and general individuals who did not come to the hospital for their own medical problems but for assisting patients such as caregivers, relatives or family members of the patients. The nonclinical group was regarded as representatives of the general population or reference group. They totaled 880 participants (58.8% females; mean age = 24.59 years old, S.D. = 9.7), min-max = 18–87 years old.

2.1.2. Clinical psychiatric participants

2.1.2.1. Outpatient psychiatric participants. This sample comprised psychiatric patients with various diagnoses based on DSM-IV-TR and DSM-5. The outpatient department accepted 40 to 50 patients for 5 to 6 psychiatrists and psychiatry residents daily. Ten percent comprised new patients. Almost all patients received medication for treatment combined with brief counseling regarding management of adverse effects of

medication and moral support. Each subject received 30 min per visit on average for medication counseling. The total sample size was 341.

2.1.2.2. Psychotherapy participants. This treatment was provided additionally for patients not recovering from regular treatment mentioned above. In this setting, psychodynamic psychotherapy model was predominantly provided by trained psychotherapists and psychiatry residents under supervision. The total sample size was 81. Of 422 subjects, 56.6% were females, mean age was 36.30 (S.D. = 14.1), and min-max = 18–88 years old. In the clinical psychiatric group, the patients were

assessed at baseline and at three months follow-up. The OI-21 scores were compared to evaluate its sensitivity to change.

2.2. Procedure

In the first stage, the scale development, a convenience sample of 150 mixed ambulatory psychiatric patients and university students were asked to complete the OI-21. In the second stage, the psychometric properties of the OI-21 were analyzed with 808 nonclinical and 422 clinical subjects, i.e., internal consistency, test-retest reliability,

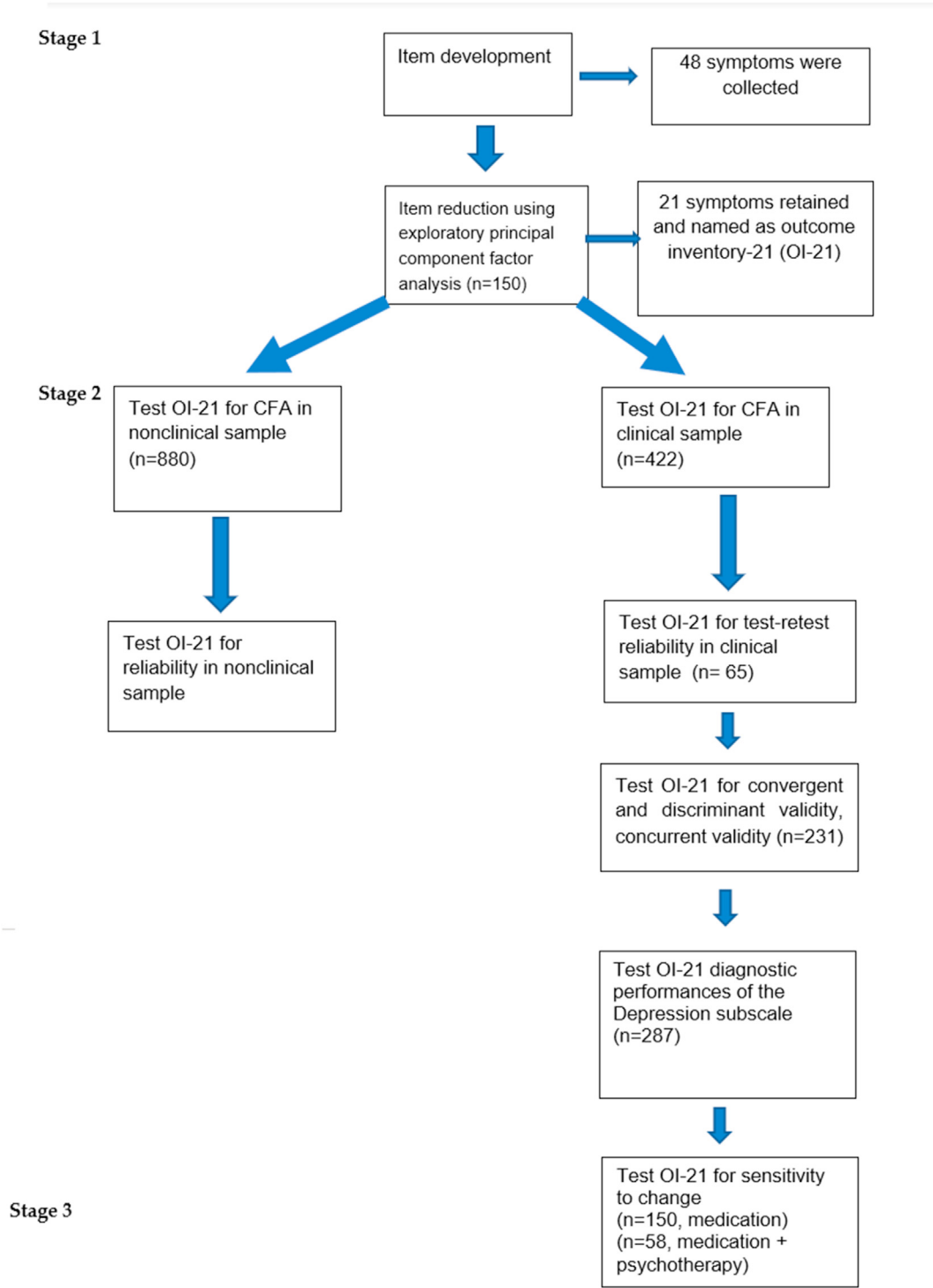


Figure 1. Title: Flowchart of the study.

confirmatory factor analysis, concurrent, convergent, discriminant validity, and diagnostic performance of a depression subscale for major depression. In the last stage, responsiveness was investigated with the 208 out of 422 clinical subjects (150 for the medication treatment group and 58 for the medication + psychotherapy group) in a three-month follow-up. Details are shown in [Figure 1](#).

The study was approved by the Ethics Committee of the Faculty of Medicine, Chiang Mai University. The study code was PSY- 458/2559 and date of approval was November 23, 2017.

2.3. Instruments

2.3.1. Development of the instrument

A draft version of the OI-21 included the common problems found in routine clinical practice. Item pooled were from the investigators' consensus as well as the investigators' research (IIP and SCL). The study aimed to create a list of items related to four main symptoms including symptoms of anxiety, depression, somatization and interpersonal difficulties. The initial scale consisted of 48 items, chosen from the high loading and bias-free items reported from various self-report measurements, with five response options, representing four domains. In item writing and selection procedures, the items should be simple, unequivocal, and understandable to the lay person regardless of education level. Therefore, no reversed item was chosen. In determining the recall period, we used a one-week period to recall the symptoms because we would like the OI to capture the symptom change on a weekly basis corresponding to usual practices in psychotherapy. The shorter period will be advantageous over a longer period such as four weeks when applied to a shorter visit than four weeks. In addition, it should be easier for the respondents to recall their symptoms within one week's time frame rather than over one week, consistent with other validated measurement such as BASIS-24 ([Cameron et al., 2007](#)), CORE-OM([Evans, 2000](#)), DASS([Lovibond and Lovibond, 1995](#)), Health Survey Short Form-36 (SF-36) ([Ware Jr, 1999](#)), OQ-45 ([Lambert et al., 2004](#)), SCL-90-R (L. [Derogatis et al., 1973](#)), and BSI(L. R. [Derogatis and Fitzpatrick, 2004](#)). Additionally, we reviewed the Patient Reported Outcome Measurement Information System (PROMIS) ([Tarescavage and Ben-Porath, 2014](#)).

2.3.2. Other measurements

2.3.2.1. The Perceived Stress Scale (PSS). The Perceived Stress Scale measures the perception of stress in the last month with ten items using a 5-point Likert scale (0 = never to 4 = very often). The total score ranges from 0 to 40, with higher scores indicating a greater perception of stress. Six questions pertain to stress and four questions pertain to control. The Thai PSS-10 showed good internal validity with an overall α of 0.85) ([Wongpakaran and Wongpakaran, 2010](#)). In this study, Cronbach's alpha was 0.84.

2.3.2.2. The multidimensional scale of perceived social support (MSPSS). The Revised Multidimensional Scale of Perceived Social Support measures a person's perception of social support ([Wongpakaran and Wongpakaran, 2012](#); [Zimet et al., 1990](#)). The MSPSS assesses three sources of social support, i.e., family members, friends, and significant others. High levels of social support are associated with low depression and anxiety symptoms. The tool has 12 questions with a seven-point Likert scale ranging from 1 = very strongly disagree to 7 = very strongly agree. Total points range between 12 and 84 points. Higher scores indicate more social support. The Cronbach alpha was 0.92 among university medical students ([Wongpakaran and Wongpakaran, 2012](#)). In this study, Cronbach's alpha was 0.92.

2.3.2.3. The experience of close relationship questionnaire-revised (ECR-R). This self-rating tool measures attachment-related anxiety and avoidance among adults ([Fraley et al., 2000](#)). The 18-item version with

five-point Likert scale response has been widely used with excellent internal consistency ([Wongpakaran and Wongpakaran, 2012](#)). Anxiety and avoidance are assessed with nine questions each, the avoidance questions are reversed, and the mean totals of the subscales are used as measures. In this study, Cronbach's alpha was 0.85 for attachment anxiety, and Cronbach's alpha of 0.844 for attachment avoidance.

2.3.2.4. The inventory of interpersonal Problems-32 (IIP-32). The IIP-32 is a self-reporting questionnaire that measures interpersonal difficulties. It asks respondents to respond to the items that they feel 'hard to do' or 'do too much'. The IIP-32 has eight subscales, i.e., domineering, vindictive, cold, nonassertive, socially inhibited, overly accommodating, self-sacrificing, and intrusive/needy. The IIP response is a five-level Likert scale, ranging from not at all (0) to extremely (4) ([Horowitz et al., 2000](#)). The Thai version of the IIP-32 demonstrated a good overall internal consistency of $\alpha = 0.95$, and 2- factor structure model with good concurrent validity ([Wongpakaran et al., 2012](#)). In this study, Cronbach's alpha was 0.89 for the whole scale, and ranged from 0.70 to 0.84 for each subscale.

2.3.2.5. The rosenberg self-esteem scale. This self-esteem tool is widely used worldwide and was developed in 1965. It is a ten-item tool that measures self-esteem with a four-point Likert scale with total scores ranging from 10 to 40. Higher scores are associated with higher levels of self-esteem. The Thai version exhibits good validity and was used in a study of university students and clinical patients with a Cronbach's alpha of 0.86 ([Wongpakaran et al., 2012](#)). In this study, Cronbach's alpha was 0.85.

2.4. Statistical analysis

Descriptive statistics was used for demographic data as well as for data screening analysis for factor analysis. Item responses showed all items had acceptable skewness and kurtosis ($< \pm 2$) ([Godin et al., 2008](#)). Exploratory factor analysis, the principal component method, was performed to determine the suitable items to be retained for the scale. Factor loading values > 0.40 were used as cut-off values for the respective item to be kept in each subscale. Confirmatory factor analysis (CFA) was conducted to test the relations between observed variables and latent variables or factors. Robust weighted least square means and variance adjusted (WLSMV) were employed for estimators as data were ordinals ([Li, 2016](#)). Regarding the fit indexes, a Comparative Fit Index (CFI) and a NonNormed Fit Index (NFI) or Tucker-Lewis Index (TLI) > 0.95 indicates good model fit, a standardized root mean square residual (SRMR), and a root-mean-square error of approximation (RMSEA) ≤ 0.06 , indicated a good as well as a χ^2/df result of < 3 ([Glader et al., 2002](#); [Gladstone et al., 2001](#); [Glassman et al., 2007](#); [Godin et al., 2008](#)). In addition, the χ^2 statistic has been used to test the goodness of model fit and nonsignificant p values associated to the test statistics, indicating that the null hypothesis of perfect fit cannot be rejected. Missing data were replaced using multiple imputation. Modification indices were adopted after the initial analysis. Mplus 7.4 was used for CFA ([Muthen & Muthen, 1998–2015](#)).

The bifactor model was analyzed because it allowed researchers to hold the idea of unidimensionality or a single common construct of OI-21 while also recognizing multidimensionality of four subscales. Model testing were examined in a series, starting from a one-dimensional model to a bifactor model, which consisted of a global severity dimension in which each item is loaded; four symptom-specific factors in which only the problem specific items are loaded without correlations between specific factors ([Brunner et al., 2012](#); [Reise et al., 2007](#)). To test for diagnostic performances of the depression subscale, receiver operating characteristics (ROC) analysis was conducted for the sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV). The diagnoses given by the psychiatrists were considered gold

standard for determining cut-off scores. The Youden Index J was used to indicate the point where the sensitivity and specificity represented the highest value. MedCalc, Version 20.010 was used for the ROC analysis (MedCalc Software, Mariakerke, Belgium).

Responsiveness or sensitivity to change (Terwee et al., 2007) was assessed by comparing the scores at baseline and at follow-up periods using the standardized response mean (SRM). SRM is the preferred value to use in comparing paired data measurements made at different time points. It is calculated by dividing the observed mean change by the standard deviation of the observed change (Stratford et al., 1996). Positive values echo improvements in the score differences. Values of 0.80, 0.50, and 0.20 were demonstrated as large, moderate, and small, respectively (Guyatt et al., 2002).

3. Results

The draft version of OI was tested in a pilot study with 150 participants from primary care settings, 57.8% females, mean age 32.78 (S.D. = 14.1). Exploratory factor analysis, Principal Axis Factoring, was used to reduce the items. It initially yielded nine factors with eigen values ranging from 13.10 to 1.04. The factors were finalized to 4, with item factor loading ≥ 0.40 retained in the scale. Finally, only 21 items remained for the four-factor OI-21.

3.1. The final version of the outcome inventory (OI-21)

The 21-item OI consists of anxiety, depression, interpersonal difficulties, and somatization subscales. The anxiety subscale had six items including items 3, 7, 9, 11, 15, and 20, The depression subscale comprised five items, including items 2, 5, 14, 18 and 21. The interpersonal difficulty subscale consisted of four items, including 4, 10, 13, and 16 and the somatization subscale comprised six items, including 1, 6, 8, 12, 17, and 19. All of which were based on a five-point Likert scale, including values of 1 (never), 2 (rarely), 3 (sometimes), 4 (frequently) and 5 (almost always). The scores range from 0 to 48, the higher the score, the higher the level of psychopathology. The instruction of the OI-21 is as follows, "For the last week - including today, please describe your feelings in response to the statements, in terms of how often you experienced them (circle the number matching your feeling) (see Appendix).

The OI-21 was then tested in a sample of 1302. All subjects' characteristics are shown in Table 1. Most participants were female (58.1%), single (60.88%), and average age was 28.39. Most patients attended to at the hospital involved depressive disorder (45.3%). The mean scores of OI-21 were significantly higher among the clinical participants than nonclinical participants. Regarding item characteristics among all 21 items, the minimum to maximum value was between 0 and 4, the means items ranged from 0.20 to 1.65, median from 0 to 2, variance from 0.41 to 1.10, skewness from -0.57 to 2.47, and kurtosis from -0.57 to 3.70.

Table 2 shows the Cronbach's alpha values of the OI-21 and the subscales for the whole, clinical and nonclinical subjects and the Cronbach's alpha values for the subscale if that respective item was deleted. Overall, Cronbach's alpha values were good to excellent, ranging from 0.77 to 0.87 for the subscales. Confirmatory factor analysis established that each item had sufficient factor loadings (estimated coefficients) on the designated factor. All factor loading coefficients were significant and ranged from 0.52 to 0.91, and higher in clinical than nonclinical subjects.

To examine which model fitted the data the best, four models were compared. Based on the fit indices shown, the bifactor model showed the best fitted model, implying that the OI-21 is sufficiently unidimensional (Table 3).

Table 4 shows the correlations between the OI-21, IIP-32, and ECR-R conducted among 150 clinical participants. Anxiety subscale score had higher magnitude of correlation with submissive domain than dominating domain, and higher with attachment anxiety than with

Table 1. Characteristics of sociodemographic data of all participants.

	All (n = 1302)	Nonclinical group (n = 880)	Clinical group (n = 422)
	n (%) or M±SD	n (%) or M±SD	n (%) or M±SD
Sociodemographic			
Sex, female n (%)	756 (58.1)	517 (58.8%)	290 (68.7)
Age	28.39 (12.56)	24.59 (9.7)	45.09 (13.4)
Years of education	12.72 (2.99)	14.58 (1.28)	10.86 (4.7)
Marital status			
Single	793 (60.88)	642 (72.92)	150 (35.6)
Lived together	355 (27.27)	167 (18.99)	188 (44.5)
Divorced/widowed	154 (11.85)	71 (8.10)	83 (19.7)
Clinical diagnosis			
Depressive disorder	-	-	191 (45.3)
Panic disorder	-	-	80 (18.1)
Alcohol abuse	-	-	46 (10.3)
Bipolar disorder	-	-	42 (9.5)
Schizophrenia and psychotic disorders	-	-	19 (4.3)
Mixed anxiety and depressive disorder	-	-	15 (3.4)
GAD	-	-	11 (2.6)
Substance abuse	-	-	8 (1.7)
Persistent depressive disorder	-	-	4 (0.9)
OCD	-	-	4 (0.9)
Other, e.g., PTSD	-	-	23 (5.2)
Outcome inventory			
Total	21.74 (12.02)	19.94 (10.27)	25.49 (14.35)
Subscale			
Anxiety	8.02 (4.50)	7.69 (4.05)	8.70 (5.2)
Depression	3.51 (3.58)	2.98 (2.94)	4.62 (4.44)
Interpersonal difficulties	4.60 (2.64)	4.28 (2.39)	5.27 (3.01)
Somatization	5.61 (4.02)	4.99 (3.50)	6.91 (4.67)

attachment avoidance. Interpersonal difficulties were, as expected, significantly related to social inhibition, and attachment avoidance. Likewise, depression was, as expected, related higher to social inhibition than other IIP subscale scores. Somatization, as anticipated, related more to attachment anxiety only than to attachment avoidance.

Table 5 shows the correlations between the OI-21 and PSS, RSES, and MSPSS scores conducted among 150 clinical participants. They all, as anticipated, showed significantly positive correlations with perceived stress but negative correlations with self-esteem and perceived social support.

Test-retest reliability was conducted among 65 clinical participants. The intraclass correlation coefficients between time 1 and time 2 for overall score, anxiety, depression, interpersonal difficulties, and somatization were: 0.80, 0.76, 0.81, 0.82, 0.83, respectively (all p < .01).

3.2. ROC analysis

Regarding accuracy of the OI-depression (OI-Dep) subscale in predicting major depression against the gold standard clinical interview diagnosis, we used the area under the ROC curve as the criterion to compare the set of items. The analysis was conducted among 287 clinical participants. The OI-Dep subscale provided area under the

Table 2. Cronbach's alpha and factor loading.

Item no.	Description	Whole sample (n = 1302)				Clinical (n = 422)				Nonclinical (n = 880)			
		α	Estimate	S.E.	Est./S.E.	α	Estimate	S.E.	Est./S.E.	α	Estimate	S.E.	Est./S.E.
ANX		0.83				0.85				0.81			
OI3	get bored	0.81	0.74	0.02	47.60	0.83	0.78	0.02	32.04	0.78	0.72	0.02	34.31
OI7	feel pressured	0.81	0.68	0.02	39.67	0.84	0.70	0.03	25.69	0.79	0.65	0.02	28.56
OI9	fear of things	0.80	0.71	0.02	41.47	0.83	0.74	0.03	26.69	0.78	0.70	0.02	31.98
OI11	poor concentration	0.80	0.68	0.02	40.40	0.82	0.73	0.03	27.82	0.77	0.65	0.02	29.10
OI15	worry	0.79	0.73	0.02	46.73	0.82	0.76	0.03	28.66	0.76	0.72	0.02	39.80
OI20	cannot work as usual	0.80	0.72	0.02	44.57	0.83	0.76	0.03	30.46	0.78	0.69	0.02	32.23
DEP		0.83				0.87				0.77			
OI2	not have happy life	0.78	0.78	0.02	48.71	0.83	0.83	0.02	38.23	0.72	0.72	0.03	29.09
OI5	feel hopeless	0.76	0.82	0.01	62.45	0.82	0.83	0.02	39.51	0.68	0.80	0.02	44.56
OI14	have no goals	0.80	0.72	0.02	39.85	0.85	0.77	0.03	27.95	0.72	0.68	0.02	28.17
OI18	feel depressed	0.79	0.85	0.01	71.90	0.83	0.87	0.02	49.91	0.72	0.83	0.02	48.61
OI21	suicidal idea	0.82	0.79	0.03	30.41	0.85	0.83	0.03	27.67	0.77	0.68	0.05	14.27
INTER		0.83				0.86				0.79			
OI4	difficult socializing	0.71	0.77	0.01	53.59	0.79	0.85	0.02	51.63	0.72	0.73	0.02	36.41
OI10	not get along	0.64	0.89	0.01	65.06	0.82	0.91	0.02	53.17	0.76	0.94	0.02	49.73
OI13	uncomfortable with people	0.63	0.75	0.02	44.24	0.86	0.80	0.02	33.09	0.74	0.69	0.02	31.50
OI16	like to be alone	0.64	0.76	0.02	48.99	0.83	0.86	0.02	44.91	0.75	0.67	0.02	30.56
SOMA		0.81				0.83				0.77			
OI1	physical pain	0.80	0.51	0.03	20.67	0.83	0.52	0.04	12.26	0.76	0.48	0.03	15.28
OI6	feel discomfort	0.76	0.72	0.02	37.66	0.80	0.77	0.03	24.82	0.71	0.68	0.02	27.93
OI8	feel numbness	0.78	0.76	0.02	38.73	0.80	0.79	0.03	26.59	0.73	0.72	0.03	26.31
OI12	headaches	0.78	0.70	0.02	37.10	0.80	0.76	0.03	24.65	0.74	0.62	0.03	24.54
OI17	shivers	0.76	0.85	0.02	50.05	0.80	0.81	0.03	30.18	0.72	0.86	0.02	37.76
OI19	sound in ears	0.78	0.78	0.02	36.20	0.81	0.78	0.03	24.33	0.75	0.76	0.03	25.67
All items		0.92				0.93				0.90			

S.E. = standard error.

Table 3. Model comparison among 1-factor, 4-factor, and bifactor model.

Model	Chi-square	df	Ch/df	CFI	TLI	RMSEA (90%CI)	SRMR
One-factor	4360.98	189	23.07	0.82	0.80	0.130 (0.127–0.134)	0.081
Four-factor	1342.63	183	7.34	0.95	0.94	0.070 (0.066–0.073)	0.044
Second order	1282.09	185	6.93	0.95	0.95	0.067 (0.064–0.071)	0.044
Bifactor	1044.01	168	6.21	0.96	0.95	0.063 (0.060–0.067)	0.038

Ch = Chi-square, df = degree of freedom, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = root-mean-square error of approximation, SRMR = standardized root mean square residual.

Table 4. Convergent and discriminant validity of the OI-21.

	IIP								ECR-R	
	DO	VI	CO	SI	NO	OA	SS	IN	AN	AV
Anxiety	.31**	.33**	.42**	.41**	.49**	.44**	.35**	.32**	.43**	.060
Depression	.10	.09	.12	.18**	.09	.08	.10	.08	.39**	.18**
Interpersonal difficulties	.16*	.39**	.48**	.60**	.11	.05	-.01	-.10	.21**	.41**
Somatization	.20**	.17**	.23**	.12	.19**	.20**	.24**	.21**	.28**	.01

*p < .05, **p < .01, IIP = the inventory of interpersonal problems, DO = domineering, VI = vindictive, CO = cold, SI = socially inhibited, NO = nonassertive, OA = overly accommodating, SS = self-sacrificing, and IN = intrusive/needy, AN = attachment anxiety, AV = attachment avoidance.

curve of 0.89 (standard error = 0.02, p < .0001) denoting good accuracy performance (Somoza et al., 1989). With the cut-off score ≥7, Youden Index J yielded 0.66 with sensitivity of 86.15%, specificity of 80.25%, positive predictive value of 78.30%, and negative predictive value of 87.50% (See Figure 2).

3.3. Responsiveness

In comparing the scores at baseline and at 3-month follow-up among 208 patients (150 in the medication group and 58 in the medication + psychotherapy group), the SRM was significant for the subscale scores

Table 5. Concurrent validity with other measurements.

	PSS	RSES	MSPSS
Anxiety	.67**	-.52**	-.31**
Depression	.53**	-.38**	-.34**
Interpersonal difficulties	.33**	-.38**	-.53**
Somatization	.34**	-.19**	-.11
Overall	.69**	-.60**	-.47**

*p < .05, **p < .01, PSS = perceived stress, RSES = Rosenberg's Self-esteem, MSPSS = Overall perceived social support.

Table 6. Comparison of responsiveness parameters after 3-month treatment.

Medication (n = 150)	Mean ± SD at baseline	Mean ± SD at Follow-up	Mean difference	SD difference	SRM (95% CI)
Anxiety	10.12 ± 5.4	9.22 ± 4.8	0.91	5.2	0.20 (0.06–0.29)
Depression	6.68 ± 5.19	5.28 ± 4.2	1.40	4.7	0.34 (0.26–0.37)
Interpersonal	5.41 ± 2.90	5.00 ± 2.8	0.41	2.9	0.15 (-0.01 to 0.27)
Somatization	12.52 ± 7.05	7.24 ± 5.9	3.35	6.5	0.38 (0.31–0.42)
Total score	38.79 ± 16.12	33.03 ± 15.0	5.75	16.8	0.34 (0.25–0.39)
Medication + psychotherapy (n = 58)					
Anxiety	14.11 ± 5.60	11.88 ± 6.5	2.23	6.1	0.34 (0.17–0.40)
Depression	9.07 ± 5.48	7.21 ± 5.5	1.86	5.5	0.31 (0.07–0.40)
Interpersonal	7.96 ± 4.08	6.55 ± 4.4	1.42	4.3	0.29 (0.09–0.38)
Somatization	8.89 ± 4.69	6.67 ± 5.0	2.23	4.9	0.36 (0.16–0.42)
Total score	40.05 ± 15.52	33.00 ± 17.9	7.05	17.2	0.41 (0.08–0.74)

SD = standard deviation, SRM = standardized response mean, CI = confidence interval.

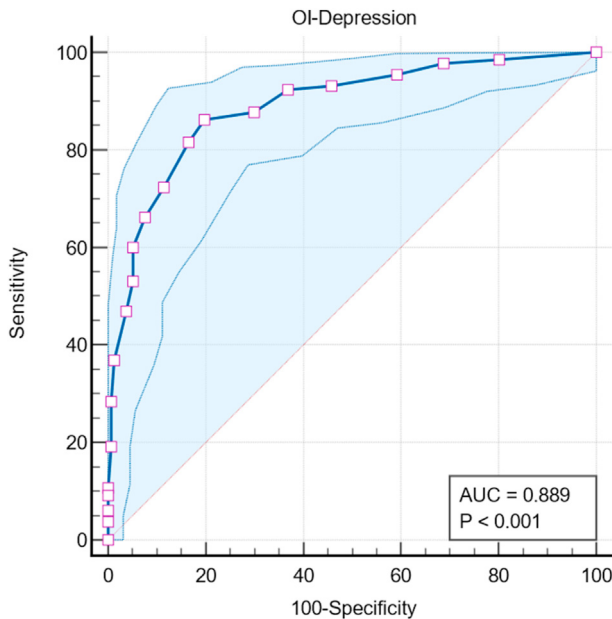


Figure 2. Title: Diagnostic performance of OI-Dep illustrating ROC curve with 95% confidence bounds. Legend: AUC = Area under the ROC curve.

(except for interpersonal difficulties) in the medication group, compared with the SRM in medication + psychotherapy group. The SRM values generally exhibited a small to moderate size of effect (Table 6).

4. Discussion

The primary goal of this study was to demonstrate the psychometric property of the newly developed self-report measure for psychopathology. Using multiple subjects rendered the OI-21 to be tested comprehensively. Regarding the construct validity, the OI-21 provided four subscales as intended with acceptable internal consistency among all subjects. Each subscale was demonstrated to have excellent internal consistency suggesting that they can be used separately. Also, sum or total scores can be used to represent the construct of overall psychological distress or symptoms. The OI-21 shows sufficient unidimensionality denoted by the fit of bifactor model to the data, which is consistent with similar measurements such as the symptom checklist-90 and the basic symptom inventory (Urbán et al., 2014). The OI-21 conforms with other outcome measurements such as OQ-45, SCL-90-R, and BSI in that the bifactor model fit the data the best (Bludworth et al., 2010; Urbán et al., 2014) and the global symptom severity explains the large correlations between symptom factors.

Regarding the factor structure, the OI-21 demonstrated a good fit for the four-factor solution model as intended. Convergent and discriminant validity revealed the interpersonal difficulties subscale had higher correlation with unfriendly (hostile) type than friendly type. Likewise, the depression subscale had higher correlation with social inhibition, which

is common and related to many psychopathologies such as depression, internet addiction, and substance abuse (Harrison et al., 2017; Husson et al., 2017; Monacis et al., 2017). The same is true for the relationship between OI subscales and attachment anxiety and avoidance, where the anxiety subscale was significantly related to attachment anxiety but not to avoidance. Vice versa, the interpersonal subscale had higher correlation with attachment avoidance than attachment anxiety. All these relationships also indicated the construct validity of the OI-21.

As expected, the validity and reliability of the OI-21 among clinical subjects were superior to nonclinical subjects because the variance of symptoms was higher among clinical than nonclinical subjects. It has been suggested that the OI-21 may be more suitable for clinical settings than nonclinical settings. Compared with other outcome measurements, the OI-21 provides good to excellent internal consistency coefficients and retest reliability as other outcome measurements, e.g., OQ-45, DASS-21, BSI, SQ-48 (Blais et al., 2015; I. V. Carlier et al., 2017; McCrae et al., 2011).

The uniqueness of the OI-21 results from adding interpersonal difficulties in the scale. It constitutes a slow to change variable compared with the other three variables. Interpersonal difficulties subscale was demonstrated convergent and discriminant validity with the interpersonal problems subscale assessed by the IIP. The interpersonal distress of the OQ-45 is comparable, demonstrating the convergent but not discriminant validity for unique interpersonal distress (Hess et al., 2010).

The OI-21 demonstrated concurrent validity with other positive and negative outcome measurements in that it, as expected, was positively related to perceived stress, but negatively related to self-esteem and perceived social support. It has been suggested that individuals with high OI scores especially anxiety and depression tend to experience high levels of stress, whereas low levels of self-esteem and perceived social support. Predictive ability should be further investigated to see how well the OI-21 can forecast such positive and negative mental health outcomes.

The OI-21 can also be used as a screening tool to identify potential clinical disorders especially major depressive disorders. The OI-Dep subscale provides preliminary data on cut-off scores for clinical subjects using the clinical diagnosis performed by psychiatrists. Compared with other measurements such as the OQ-45 that yielded AUC values of

0.77–0.85 for symptoms, interpersonal, and social role subscales (Iraurgi and Penas, 2021), OI-Dep illustrated good performance as it yielded an AUC of 0.89 with the cut-off score of 7. However, the optimal cut-off score may depend on the prevalence and the setting of the study.

Sensitivity to change is one of the important attributes of the outcome measurements. Compared with other validated measurements such as the BASIS-24, CORE-OM, OQ-45, and PROMIS (I. V. Carlier et al., 2017; Errázuriz et al., 2017), the OI-21 demonstrated its sensitivity to change in both total score, and subscale scores. The relatively low effect size may be due to the interval of the follow-up period. The fact that the interpersonal difficulties are not as sensitive to change as anxiety, depression, and the somatization subscale in the medication group may be because this subscale tends to more like a “trait” than a “state” like symptoms (Denollet and Pedersen, 2008). They are slow to change. The results showed that the group receiving additional psychotherapy appeared to have more interpersonal changes in addition to symptoms. This suggests that the interpersonal difficulties subscale may be a suitable indicator to measure a slow to change outcome that normally requires intensive treatment such as treatment combining medication and psychotherapy (Schauenburg et al., 2000).

Some clinicians may be interested in finding a cut-off score to determine change for each symptom in a single subject. The cut-off score can be obtained by multiplying 1.96 by the standard error of change calculated by “initial standard deviation *sqrt (2)*sqrt (1-reliability)” (Hageman and Arrindell, 1993). The cut-off score for clinical and nonclinical subjects; should however, be calculated separately. For example, the standard deviation of anxiety score in the nonclinical sample was 4.05, and the Cronbach's alpha was 0.81; therefore, the standard error of change given these values was 2.52. The reliable change criterion was 4.93 (1.96 × 2.52). For clinical subjects, the standard deviation of anxiety score in the nonclinical sample was 5.20, and the Cronbach's alpha was 0.85; therefore, the standard error of change given these values was 2.82. The reliable change criterion was 5.53 (1.96 × 2.82). The change of anxiety scores should be greater than 4.93 and 5.53 for nonclinical and clinical subjects, respectively to be considered reliable change.

During this COVID-19 pandemic, research demonstrated an increased level of mental health problems, e.g., anxiety, depression and somatization of people across population subgroups, especially in nonclinical population (Dragioti et al., 2021; Hu et al., 2022; Zhang et al., 2022). The level of some symptoms, depression and loneliness due to COVID, were raised in nonclinical populations to the extent that nondifferences between clinical and nonclinical subjects were observed (Rek et al., 2022). The role of self-report questionnaires, including OI-21, should be suitable for identifying and monitoring such mental health problems.

4.1. Limitations of the study and future research

The study was conducted in one setting using nonclinical participants, i.e., university students, and general individuals who did not come to the hospital for their own medical problems but assist patients such as caregivers, relatives, or family members of the patients. However, they may not constitute actual representatives of a general population. Study in a general population should be encouraged. The same is true for the clinical sample, which was limited to outpatient psychiatric and psychotherapy clinics, whose characteristics may not be generalized to other clinical settings. More research should be encouraged especially in general practice, and medical, counseling clinics. In addition, the study used convenient sampling that may be biased in terms of sampling error. A more systemic randomization fashion would ensure more reliable score values when calculating the reliable change index. In terms of psychometric properties, measurement invariance or differential item functioning, e.g., bias due to sex, age, or education level, should be further investigated especially using modern test method, i.e., item response theory.

5. Conclusion

The OI-21 is a reliable and valid instrument providing a comprehensive survey of psychological distress. It has substantial satisfactory psychometric properties; and therefore, can be used in clinical, research and service settings. The questionnaire is not a diagnostic aid but can be very useful in the clinical monitoring of pathologies (both in pharmacologic and psychotherapeutic terms) and follow-up.

The common symptoms such as anxiety, depression, and somatization can be used for assessing symptom outcomes as well as for monitoring routine psychiatric services that rely mainly on psychotropic medication. Also, the interpersonal difficulties subscale can be used for a more rigorous treatment, i.e., combined medication and psychotherapy. The depression subscale of the OI-21 can also be used as a screening instrument for major depression. Further testing of the use and validity of the OI-21 in other clinical or cultural settings is encouraged.

Institutional review Board statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional review Board (or Ethics Committee) of the Faculty of Medicine, Chiang Mai University (study code, PSY- 458/2559 and date of approval, November 23, 2017).

Informed consent statement

All patients provided written informed consent to the study.

Declarations

Author contribution statement

Nahathai Wongpakaran: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Tinakon Wongpakaran: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Zsuzsanna KÖVI: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by the Faculty of Medicine Research Fund of Chiang Mai University (grant no. 094/2560).

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2022.e09682>.

Acknowledgements

We are thankful to all participants and our research assistants involving the data collecting process.

Appendix

No.	Outcome Inventory-21	Sex	Female	Male	Age	years
	For the last week - including today, please describe your feelings in response to the statements, in terms of how often you experienced them (Circle the number that matches your feeling)	Never	Rarely	Occasionally	Frequently	Almost Always
	There are a total of 21 statements	1	2	3	4	5
1	I experience physical pain across many parts of my body.					
2	I believe that I cannot have a happy life - as others do.					
3	I get bored with things easily.					
4	I find it difficult to get to know other people.					
5	I feel hopeless about my life.					
6	I feel discomfort in my head and/or nose.					
7	I feel pressured by the people or things around me.					
8	I feel numbness or a tickling sensation.					
9	I feel unhappy due to fear of specific things or situations.					
10	I do not get along with others.					
11	I am unable to concentrate while performing tasks.					
12	I experience headaches.					
13	I feel uncomfortable with people that are not family members.					
14	I feel I have no goals in my life.					
15	I worry about almost everything.					
16	I like to be alone instead of being social.					
17	I experience the shivers.					
18	I feel depressed.					
19	I hear a ringing/humming sound in my ears.					
20	I cannot work or study as well as I should.					
21	I have suicidal ideas.					

References

Allen, L.A., Woolfolk, R.L., Escobar, J.I., Gara, M.A., Hamer, R.M., 2006. Cognitive-behavioral therapy for somatization disorder: a randomized controlled trial. *Arch. Intern. Med.* 166 (14), 1512–1518.

American Psychiatric, A., 1996. *DSM-IV Sourcebook*. American Psychiatric Press, Washington.

APA, 2013. *Diagnostic and Statistical Manual of Mental Disorder (DSM-5)*, fifth ed. Washington DC.

Azocar, F., Areán, P., Miranda, J., Muñoz, R.F., 2001. Differential item functioning in a Spanish translation of the Beck depression inventory. *J. Clin. Psychol.* 57 (3), 355–365.

Barkham, M., Margison, F., Leach, C., Luccock, M., Mellor-Clark, J., Evans, C., McGrath, G., 2001. Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. *Clinical Outcomes in Routine Evaluation-Outcome Measures. J. Consult. Clin. Psychol.* 69 (2), 184–196.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J.E., Erbaugh, J.K., 1961. An inventory for measuring depression. *Arch. Gen. Psychiatr.* 4.

Berghout, C.C., Zevalkink, J., Katzko, M.W., de Jong, J.T., 2012. Changes in symptoms and interpersonal problems during the first 2 years of long-term psychoanalytic psychotherapy and psychoanalysis. *Psychol. Psychother.* 85 (2), 203–219.

Blais, M., Blagys, M.D., Rivas-Vazquez, R., Bello, I., Sinclair, S.J., 2015. Development and initial validation of a brief symptom measure. *Clin. Psychol. Psychother.* 22 (3), 267–275 quiz 276-267.

Bludworth, J.L., Tracey, T.J., Glidden-Tracey, C., 2010. The bilevel structure of the Outcome Questionnaire-45. *Psychol. Assess.* 22 (2), 350–355.

Broekman, B.F., Nyunt, S.Z., Niti, M., Jin, A.Z., Ko, S.M., Kumar, R., Ng, T.P., 2008. Differential item functioning of the geriatric depression scale in an Asian population. *J. Affect. Disord.* 108 (3), 285–290.

Brunner, M., Nagy, G., Wilhelm, O., 2012. A tutorial on hierarchically structured constructs. *J. Pers.* 80 (4), 796–846.

Cameron, I.M., Cunningham, L., Crawford, J.R., Eagles, J.M., Eisen, S.V., Lawton, K., Hamilton, R.J., 2007. Psychometric properties of the BASIS-24© (Behaviour and symptom identification scale-revised) mental health outcome measure. *Int. J. Psychiatr. Clin. Pract.* 11 (1), 36–43.

Carlier, I., Schulte-Van Maaren, Y., Wardenaar, K., Giltay, E., Van Noorden, M., Vergeer, P., Zitman, F., 2012. Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. *Psychiatr. Res.* 200 (2-3), 904–910.

Carlier, I.V., Kovács, V., van Noorden, M.S., van der Feltz-Cornelis, C., Mooij, N., Schulte-van Maaren, Y.W., Giltay, E.J., 2017. Evaluating the responsiveness to therapeutic change with routine outcome monitoring: a comparison of the symptom

questionnaire-48 (SQ-48) with the brief symptom inventory (BSI) and the outcome questionnaire-45 (OQ-45). *Clin. Psychol. Psychother.* 24 (1), 61–71.

Choi, H., Fogg, L., Lee, E.E., Choi Wu, M., 2009. Evaluating differential item functioning of the CES-D scale According to caregiver status and cultural context in Korean women. *J. Am. Psychiatr. Nurses Assoc.* 15 (4), 240–248.

Denollet, J., Pedersen, S.S., 2008. Prognostic value of Type D personality compared with depressive symptoms. *Arch. Intern. Med.* 168 (4), 431–432.

Derogatis, L., Lipman, R., Covi, L., 1973. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol. Bull.* 9, 13–28.

Derogatis, L.R., Fitzpatrick, M., 2004. The SCL-90-R, the brief symptom inventory (BSI), and the BSI-18. In: *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults*, third ed.ume 3. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 1–41.

Dragioti, E., Li, H., Tsitsas, G., Lee, K.H., Choi, J., Kim, J., Solmi, M., 2021. A large-scale meta-analytic atlas of mental health problems prevalence during the COVID-19 early pandemic. *J. Med. Virol.*

Ellsworth, J.R., Lambert, M.J., Johnson, J., 2006. A comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clin. Psychol. Psychother.* 13 (6), 380–391.

Errázuriz, P., Opazo, S., Behn, A., Silva, O., Gloger, S., 2017. Spanish adaptation and validation of the outcome questionnaire OQ-30.2. *Front. Psychol.* 8, 673.

Evans, J M-C, F M, M B, K A, J C, G M C, 2000. CORE: clinical outcomes in routine evaluation. *J. Ment. Health* 9 (3), 247–255.

Forero, C.G., Adroher, N.D., Stewart-Brown, S., Castellví, P., Codony, M., Vilagut, G., Alonso, J., 2014. Differential item and test functioning methodology indicated that item response bias was not a substantial cause of country differences in mental well-being. *J. Clin. Epidemiol.* 67 (12), 1364–1374.

Fralely, R.C., Waller, N.G., Brennan, K.A., 2000. An item response theory analysis of self-report measures of adult attachment. *J. Pers. Soc. Psychol.* 78 (2), 350–365.

Giannopoulou, I., Pasalari, E., Bali, P., Grammatikaki, D., Ferentinos, P., 2021. Psychometric properties of the revised child anxiety and depression scale in Greek adolescents. *Clin. Child Psychol. Psychiatr.* 13591045211056502.

Glader, E.L., Stegmayr, B., Asplund, K., 2002. Poststroke fatigue: a 2-year follow-up study of stroke patients in Sweden. *Stroke* 33 (5), 1327–1333.

Gladstone, G.L., Mitchell, P.B., Parker, G., Wilhelm, K., Austin, M.P., Eysers, K., 2001. Indicators of suicide over 10 years in a specialist mood disorders unit sample. *J. Clin. Psychiatr.* 62 (12), 945–951.

Glassman, A.H., Bigger, J.T., Gaffney, M., Van Zyl, L.T., 2007. Heart rate variability in acute coronary syndrome patients with major depression: influence of sertraline and mood improvement. *Arch. Gen. Psychiatr.* 64 (9), 1025–1031.

Godin, O., Dufouil, C., Maillard, P., Delcroix, N., Mazoyer, B., Crivello, F., Tzourio, C., 2008. White matter lesions as a predictor of depression in the elderly: the 3C-Dijon study. *Biol. Psychiatr.* 63 (7), 663–669.

- Goldberg, D., Williams, P., 1988. *A User's Guide to the General Health Questionnaire*. NFERNELSON, Oxford.
- Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K.W., Norman, G.R., 2002. Methods to explain the clinical significance of health status measures. *Mayo Clin. Proc.* 77 (4), 371–383.
- Hageman, W.J.J.M., Arrindell, W.A., 1993. A further refinement of the reliability change (RC) index by improving the pre-post difference score: introducing RELIAB. *Behav. Res. Ther.* 31 (7), 693–700.
- Hanel, G., Henningsen, P., Herzog, W., Sauer, N., Schaefer, R., Szecsenyi, J., Löwe, B., 2009. Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. *J. Psychosom. Res.* 67 (3), 189–197.
- Harrison, A.J., Timko, C., Blonigen, D.M., 2017. Interpersonal styles, peer relationships, and outcomes in residential substance use treatment. *J. Subst. Abuse Treat.* 81, 17–24.
- Hess, T., Rohlfing, J., Hardy, A., Glidden-Tracey, C., Tracey, T., 2010. An examination of the "interpersonalness" of the outcome questionnaire. *Assessment* 17 (3), 396–399.
- Hoge, E.A., Tamrakar, S.M., Christian, K.M., Mahara, N., Nepal, M.K., Pollack, M.H., Simon, N.M., 2006. Cross-cultural differences in somatic presentation in patients with generalized anxiety disorder. *J. Nerv. Ment. Dis.* 194 (12), 962–966.
- Horowitz, L.M., Alden, L.E., Wiggins, J.S., Pincus, A.L., 2000. *IIP Inventory of Interpersonal Problems Manual*. the psychological corporation, Texas.
- Hu, N., Deng, H., Yang, H., Wang, C., Cui, Y., Chen, J., Li, Y., 2022. The pooled prevalence of the mental problems of Chinese medical staff during the COVID-19 outbreak: a meta-analysis. *J. Affect. Disord.* 303, 323–330.
- Husson, O., Denollet, J., Ezenand, N.P., Mols, F., 2017. Personality, health behaviors, and quality of life among colorectal cancer survivors: results from the PROFILES registry. *J. Psychosoc. Oncol.* 35 (1), 61–76.
- Iraurgi, I., Penas, P., 2021. Outcomes assessment in psychological treatment: Spanish adaptation of OQ-45 (outcome questionnaire). *Rev. Latinoam. Psicol.* 53, 56–63.
- Kaiser, T., Herzog, P., Voderholzer, U., Brakemeier, E.L., 2022. Out of sight, out of mind? High discrepancy between observer- and patient-reported outcome after routine inpatient treatment for depression. *J. Affect. Disord.* 300, 322–325.
- Kalibatseva, Z., Leong, F.T., Ham, E.H., 2014. A symptom profile of depression among Asian Americans: is there evidence for differential item functioning of depressive symptoms? *Psychol. Med.* 44 (12), 2567–2578.
- Kemmelmeier, M., 2016. Cultural differences in survey responding: issues and insights in the study of response biases. *Int. J. Psychol.* 51 (6), 439–444.
- Kirmayer, L.J., 2001. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *J. Clin. Psychiatr.* 62 (Suppl 13), 22–28 discussion 29–30.
- Kirmayer, L.J., Young, A., 1998. Culture and somatization: clinical, epidemiological, and ethnographic perspectives. *Psychosom. Med.* 60 (4), 420–430.
- Kohlmann, S., Gierk, B., Hilbert, A., Brähler, E., Löwe, B., 2016. The overlap of somatic, anxious and depressive syndromes: a population-based analysis. *J. Psychosom. Res.* 90, 51–56.
- Lambert, M.J., Gregersen, A.T., Burlingame, G.M., 2004. The outcome questionnaire-45. In: *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults*, third ed.ume 3. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 191–234.
- Li, C.H., 2016. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48 (3), 936–949.
- Liebherz, S., Rabung, S., 2014. Do patients' symptoms and interpersonal problems improve in psychotherapeutic hospital treatment in Germany? - a systematic review and meta-analysis. *PLoS One* 9 (8), e105329.
- Lo Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., Lambert, M.J., 2008. The factorial structure of the Outcome Questionnaire-45: a study with an Italian sample. *Clin. Psychol. Psychother.* 15 (6), 418–423.
- 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. *Behav. Res. Ther.* 33 (3), 335–343.
- Lyu, D., Wu, Z., Wang, Y., Huang, Q., Cao, T., Zhao, J., Fang, Y., 2019. Disagreement and factors between symptom on self-report and clinician rating of major depressive disorder: a report of a national survey in China. *J. Affect. Disord.* 253, 141–146.
- Löwe, B., Spitzer, R.L., Williams, J.B., Mussell, M., Schellberg, D., Kroenke, K., 2008. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen. Hosp. Psychiatr.* 30 (3), 191–199.
- McCrae, R.R., Kurtz, J.E., Yamagata, S., Terracciano, A., 2011. Internal consistency, retest reliability, and their implications for personality scale validity. *Pers. Soc. Psychol. Rev. : Off. J. Soc. Personal. Social Psychol. Inc* 15 (1), 28–50.
- Means-Christensen, A.J., Roy-Byrne, P.P., Sherbourne, C.D., Craske, M.G., Stein, M.B., 2008. Relationships among pain, anxiety, and depression in primary care. *Depress. Anxiety* 25 (7), 593–600.
- Moessner, M., Gallas, C., Haug, S., Kordy, H., 2011. The clinical psychological diagnostic system (KPD-38): sensitivity to change and validity of a self-report instrument for outcome monitoring and quality assurance. *Clin. Psychol. Psychother.* 18 (4), 331–338.
- Monacis, L., de Palo, V., Griffiths, M.D., Sinatra, M., 2017. Exploring individual differences in online addictions: the role of identity and attachment. *Int. J. Ment. Health Addiction* 15 (4), 853–868.
- Muthen, L.K., Muthen, B.O., 1998-2015. *Mplus User's Guide*, 7th Edition. Los Angeles, CA.
- Nebeker, R.S., Lambert, M.J., Huefner, J.C., 1995. Ethnic differences on the outcome questionnaire. *Psychol. Rep.* 77 (3 Pt 1), 875–879.
- Parkerson, H.A., Thibodeau, M.A., Brandt, C.P., Zvolensky, M.J., Asmundson, G.J., 2015. Cultural-based biases of the GAD-7. *J. Anxiety Disord.* 31, 38–42.
- Quilty, L.C., Mainland, B.J., McBride, C., Bagby, R.M., 2013. Interpersonal problems and impacts: further evidence for the role of interpersonal functioning in treatment outcome in major depressive disorder. *J. Affect. Disord.* 150 (2), 393–400.
- Recklitis, C.J., Rodriguez, P., 2007. Screening childhood cancer survivors with the brief symptom inventory-18: classification agreement with the symptom checklist-90-revised. *Psycho Oncol.* 16 (5), 429–436.
- Reich, H., Rief, W., Brähler, E., Mewes, R., 2018. Cross-cultural validation of the German and Turkish versions of the PHQ-9: an IRT approach. *BMC Psychol.* 6 (1), 26.
- Reise, S.P., Morizot, J., Hays, R.D., 2007. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual. Life Res.* 16 (Suppl 1), 19–31.
- Rek, S.V., Freeman, D., Reinhard, M.A., Bühner, M., Grosen, S., Falkai, P., Padberg, F., 2022. Differential psychological response to the COVID-19 pandemic in psychiatric inpatients compared to a non-clinical population from Germany. *Eur. Arch. Psychiatr. Clin. Neurosci.* 272 (1), 67–79.
- Romppel, M., Hinz, A., Finck, C., Young, J., Brähler, E., Glaesmer, H., 2017. Cross-cultural measurement invariance of the General Health Questionnaire-12 in a German and a Colombian population sample. *Int. J. Methods Psychiatr. Res.*
- Schauburg, H., Kuda, M., Sammet, I., Strack, M., 2000. The influence of interpersonal problems and symptom severity on the duration and outcome of short-term psychodynamic psychotherapy. *Psychother. Res.* 10 (2), 133–146.
- Skodol, A.E., 2011. Revision of the personality disorder model for DSM-5. *Am. J. Psychiatr.* 168 (1), 97 author reply 97–98.
- Skodol, A.E., Bender, D.S., 2009. The future of personality disorders in DSM-V? *Am. J. Psychiatr.* 166 (4), 388–391.
- Somoza, E., Soutullo-Esperon, L., Mossman, D., 1989. Evaluation and optimization of diagnostic tests using receiver operating characteristic analysis and information theory. *Int. J. Bio Med. Comput.* 24 (3), 153–189.
- Stein, M., Hilsenroth, M., Pincus, A., Primavera, L., 2009. Validity of DSM-IV axis V global assessment of relational functioning scale: a multimethod assessment. *J. Nerv. Ment. Dis.* 197 (1), 50–55.
- Stratford, P.W., Binkley, J.M., Riddle, D.L., 1996. Health status measures: strategies and analytic methods for assessing change scores. *Phys. Ther.* 76 (10), 1109–1123.
- Tarescavage, A.M., Ben-Porath, Y.S., 2014. Psychotherapeutic outcomes measures: a critical review for practitioners. *J. Clin. Psychol.* 70 (9), 808–830.
- Terluin, B., Unalan, P.C., Turfaner Sipahioglu, N., Arslan Özkul, S., van Marwijk, H.W., 2016. Cross-cultural validation of the Turkish Four-Dimensional Symptom Questionnaire (4DSQ) using differential item and test functioning (DIF and DTF) analysis. *BMC. Fam. Pract.* 17, 53.
- Terluin, B., van Marwijk, H.W.J., Adèr, H.J., de Vet, H.C.W., Penninx, B.W.J.H., Hermens, M.L.M., Stalman, W.A.B., 2006. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC. Psychiatr.* 6 (1), 34.
- Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., de Vet, H.C., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 60 (1), 34–42.
- Trull, T.J., Distel, M.A., Carpenter, R.W., 2011. DSM-5 Borderline personality disorder: at the border between a dimensional and a categorical view. *Curr. Psychiatr. Rep.* 13 (1), 43–49.
- Uebelacker, L.A., Strong, D., Weinstock, L.M., Miller, I.W., 2009. Use of item response theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. *Psychol. Med.* 39 (4), 591–601.
- Urbán, R., Kun, B., Farkas, J., Paksi, B., Kókonyei, G., Unoka, Z., Demetrovics, Z., 2014. Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. *Psychiatr. Res.* 216 (1), 146–154.
- van Dessel, N., den Boeft, M., van der Wouden, J.C., Kleinstäuber, M., Leone, S.S., Terluin, B., van Marwijk, H., 2014. Non-pharmacological interventions for somatoform disorders and medically unexplained physical symptoms (MUPS) in adults. *Cochrane Database Syst. Rev.* 11, CD011142.
- Wardenaar, K.J., van Veen, T., Giltay, E.J., de Beurs, E., Penninx, B.W., Zitman, F.G., 2010. Development and validation of a 30-item short adaptation of the mood and anxiety symptoms questionnaire (MASQ). *Psychiatr. Res.* 179 (1), 101–106.
- Ware Jr., J.E., 1999. SF-36 health survey. In: *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, second ed. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 1227–1246.
- Whipple, J.L., Lambert, M.J., 2011. Outcome measures for practice. *Annu. Rev. Clin. Psychol.* 7, 87–111.
- Wiesner, M., Chen, V., Windle, M., Elliott, M.N., Grunbaum, J.A., Kanouse, D.E., Schuster, M.A., 2010. Factor structure and psychometric properties of the Brief Symptom Inventory-18 in women: a MACS approach to testing for invariance across racial/ethnic groups. *Psychol. Assess.* 22 (4), 912–922.
- Wongpakaran, N., Wongpakaran, T., 2012. A revised Thai multi-dimensional scale of perceived social support. *Spanish J. Psychol.* 15 (3), 1503–1509.
- Wongpakaran, N., Wongpakaran, T., Kuntawong, P., 2019. Evaluating hierarchical items of the geriatric depression scale through factor analysis and item response theory. *Heliyon* 5 (8), e02300.
- Wongpakaran, N., Wongpakaran, T., van Reekum, R., 2013. Discrepancies in Cornell Scale for Depression in Dementia (CSDD) items between residents and caregivers, and the CSDD's factor structure. *Clin. Interv. Aging* 8, 641–648.
- Wongpakaran, T., Wongpakaran, N., Boripuntakul, T., 2011. Symptom checklist-90 (SCL-90) in a Thai sample. *J. Med. Assoc. Thai.* 94 (9), 1141–1149.

- Wongpakaran, T., Wongpakaran, N., Sirithethawee, U., Pratoomsri, W., Burapakajompong, N., Rangseekajee, P., Temboonkiat, A., 2012. Interpersonal problems among psychiatric outpatients and non-clinical samples. *Singap. Med. J.* 53 (7), 481–487.
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., Leirer, V.O., 1982. Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* 17 (1), 37–49.
- Zhang, S.X., Miller, S.O., Xu, W., Yin, A., Chen, B.Z., Delios, A., Chen, J., 2022. Meta-analytic evidence of depression and anxiety in Eastern Europe during the COVID-19 pandemic. *Eur. J. Psychotraumatol.* 13 (1), 2000132.
- Zimet, G.D., Powell, S.S., Farley, G.K., Werkman, S., Berkoff, K.A., 1990. Psychometric characteristics of the multidimensional scale of perceived social support. *J. Pers. Assess.* 55 (3-4), 610–617.