Psychiatr. Pol. 2015; 49(5): 1043–1070
PL ISSN 0033-2674 (PRINT), ISSN 2391-5854 (ONLINE)
www.psychiatriapolska.pl
DOI: http://dx.doi.org/10.12740/PP/59064

The orthogonal-oblique bi-level model of the Outcome Questionnaire (OQ-45.2): Polish adaptation based on factor analysis

Witold Simon[1], Piotr Śliwka[2], Jerzy A. Sobański[3], Katarzyna Klasa[4], Paweł Sala[5], Wojciech Żak[6], Gregory Busath[7], Michael J. Lambert[7]

[1]Konsonans, Psychotherapy Centre, Warsaw
[2]Department of Mathematics and Natural Sciences,
Cardinal Stefan Wyszynski University in Warsaw
[3]Department of Psychotherapy, Jagiellonian University Medical College
[4]Department of Psychotherapy, University Hospital in Krakow
[5]Private practice, Warsaw
[6] Department of Adult Psychiatry, Poznan University of Medical Sciences
[7]Psychology Department, Brigham Young University, Provo, UT, USA

Summary

Aim. The aim of this project was to conduct the cross-cultural factorial validation of the Outcome Questionnaire (OQ-45.2) using the Polish population.

Material and method. Data were obtained from day-patients (n = 211), inpatients (n = 234), outpatients (n = 137) and non-patients (n = 426). Statistical analyses included: parallel analysis, exploratory factor analysis, confirmatory factor analysis, correlation analysis, criterion equivalence, clinical significance and reliable change index, and test-retest.

Results. Statistical analyses provided the strongest support for the bi-level model of the total score, five orthogonal (subscales specific for the Polish OQ, i.e. Social Conflicts and Addictive Behaviours; and original yet modified subscales, i.e. Symptom Distress, Interpersonal Relations, and Social Role), and two oblique factors (Somatisation and Anxiety, Social Role

David M. Kennedy Center Research Grant supported this project. The authors thank those individuals who helped to collect the data: Sławomira Bering, Piotr Domaradzki, Małgorzata Dudkiewicz, Paweł Dybka, Wioletta Gajowa, Marek Gajowy, Paweł Jankowski, Andrzej Karczewski, Anna Kiełbasińska, Jan Kraszewski, Aleksandra Kulczyńska, Leszek Madeyski, Andrzej Miziołek, Magdalena Musioł, Agata Rusak, Ewelina Seroka, Piotr Szetelnicki, Janusz Sztencel, Piotr Waszkiewicz, Joanna Wergiles, Andrzej Winkler, Andrzej Zembrzuski, and Beata Zimowska.

2). The psychometric properties of the Polish OQ were found to be adequate and similar to the original American OQ and its international adaptations. Specific for the Polish OQ cut-off scores for clinical significance were established. The role of cultural differences and the passage of time in the process of the cross-cultural validation were elaborated upon.

Conclusions. The Polish version of the OQ 45-2 has been recognised as an instrument adequately measuring general functioning as well as specific areas of functioning of the individual (i.e. interpersonal relations; social role performance; social conflicts; symptom distress; somatisation and anxiety; addictive behaviours). Results of this factorial analysis seem to be valuable for both clinicians using the OQ-45.2 and for creators of any psychotherapy outcome measure.

Key words: psychotherapy effectiveness, symptom checklist, outcome questionnaire

Introduction

Efforts to enhance psychotherapy effectiveness have resulted in the development of the Outcome Questionnaire (OQ-45.2) [1]. The OQ has been recognised as the fourth most frequently used self-report measure of psychotherapy outcome [2], which probably occurred due to its following features: (1) subscale structure (Symptom Distress, Interpersonal Relations, and Social Role) allowing for measuring improvement and deterioration of various dimensions of client functioning; (2) applicability for a variety of disorders and diverse life struggles, which enables comparison of different patient groups; (3) sensitivity to change and thus ability to track patient's progress by repeated measurements on a session-by-session basis; (4) predictive ability to identify those patients who are off track for a positive outcome and thus are at risk of deterioration and/or drop-out; and (5) signal-alarm feedback system of progress graphs, implemented in the OQ-Analyst software, which provides therapists with an instant access to clinical information and thus allows them to maximise their clients' treatment outcome, i.e. to intervene before treatment failure occurs [3].

Numerous studies have been conducted validating the original version of OQ in diverse populations and treatment settings: (1) outpatients from university centre [4]; (2) outpatients from a psychiatric unit of the general hospital [5]; (3) outpatients from a community-based substance abuse centre [6]; and (4) inpatients from an eating disorders clinic [7]. Meta/mega-analytic review of the OQ [8] confirms that the OQ based progress feedback with alert signals has a statistically and clinically significant effect on outcome in reducing deterioration rates in patients predicted to be treatment failures, i.e. from the baseline of 20% in not-on-track cases to about 5.5%. Several language versions of the OQ have been adapted, e.g., Dutch [9], German [10], Italian [11], and Swedish [12].

One of the most frequently administered outcome measures in Poland is Symptom Checklist "O" (Kwestionariusz Objawowy "O" – KO "O") and its versions [13]. They were applied in psychotherapy process research [14], including projects assessing risk of therapy termination [15]; as well as in studies of patients suffering from sexual dysfunctions [16] or suicidal thoughts [17].

Aim

The current project aims to establish the cross-cultural adaptation of the Polish OQ, which implies assessing to what extent the Polish factor structure and normative scores are equivalent with the American version as well as determining the psychometric properties of the Polish OQ. We expect the results of our study to broaden the understanding of the OQ-45.2 factor structure since statistical evidence supporting the three factor structure of the OQ-45 has been widely discussed [9, 11, 12, 18–21].

Psychometric Properties of the Outcome Questionnaire (OQ-45.2)

The OQ-45.2 is a 45-item self-report measure, using a 5-point scale (0 = never, 1 = rarely, 2 = sometimes, 3 = frequently, 4 = almost always), which yields a range of total scores from 0 to 180.

In order to limit the likelihood of biased responses, 20% of items (9/45) have reverse scoring. High scores indicate greater levels of symptoms and/or poorer functioning.

Reliability of the OQ is acceptable for all subscales and the total score. Lambert et al. [22] reported internal consistency reliability (Cronbach's alpha) for the OQ-45 of 0.93 and a 3-week test–retest reliability value of 0.84 for the OQ-45 total score. Concurrent validity of the OQ-45 total score has been examined by correlating it with the Symptom Checklist-90 [23], Beck Depression Inventory [24], and the State-Trait Anxiety Inventory [25]. All of the concurrent validity figures with the OQ-45 and these instruments were significant at the 0.01 level with a range of rs from 0.50 to 0.85. The score of 64 has been suggested as a cut-off point for differentiation between a nonclinical population of non-patients and a clinical population of patients; and a difference of 14 points has been indicated as a reliable change [22].

Factor Structure

In addition to the total score, the OQ-45 has three orthogonal subscales attempting to measure separate domains of functioning. The Interpersonal Relations subscale (IR, 11 items) assesses the quality of interpersonal relations e.g., in terms of conflict intensity or isolation tendencies. The Social Role subscale (SR, 9 items) addresses occupational and leisure activities. The Symptom Distress subscale (SD, 25 items) evaluates various symptoms, mainly depression and anxiety. It is important to state, that the three factor structure has not emerged as a result of exploratory factor analysis, but has been established rather in an arbitrary way by experienced clinicians. Having conducted a thorough research of literature review they rationally selected items for each subscale, and subsequently performed the inter-item correlations [1].

Consequently, even these authors, who to a certain extent confirm the adequacy of the three subscale model, indicate that other solutions perhaps more accurately capture the multidimensionality of the OQ. Sometimes, when the three factor model was acknowledged as a reasonable fit, as in de Jong et al. [9] study of the Dutch population, still some additional factors, e.g. the Anxiety and Somatic Distress, were proposed as a valuable account for the unexplained variance. Similarly authors of the Swedish version of the OQ [12] suggested a ten factor model as a potentially more adequate structure of the OQ.

Frequently, when sufficient support for the three factor structure was not found, the non-orthogonal exploratory factor analysis pointed to a general oblique factor model, in addition to the SD, the SR, and the IR subscales. For example, although Bludworth et al. [18] found the three factor model somewhat fitting their data, their results indicated the four factor bi-level model, with a non-overlapping General Distress factor, as a more adequate structure. Analogically, Lo Coco et al. [11] proposed the four factor bi-level solution with the Overall Maladjustment factor representing one's general functioning across the domains. These findings reflect earlier results of Mueller et al. [21] who pointed out the overlap between the three original subscales, which in their opinion may substantially cloud the dimensional interpretation of the OQ-45.

Often when fit for the three factor model was found poor, the number of orthogonal subscales substantially increased. For example, eight independent factors were indicated by Minami et al [20]: Anhedonia, Psychological Distress, Physical Distress, Loss of Productivity, Lack of Intimacy, Problematic Substance Use, Interpersonal Conflict, and Stress.

Alternatively, when a confirmatory factor analysis did not prove the three factor model as a satisfactory enough fit for the data, not only the structure of the OQ was modified, but also the number of items was reduced. Kim et al. [19] reduced the number of items to 26 and found the four factor model (Anxiety-Somatisation, Negative Self-Worth, Substance Use, Loss of Interest) as an improvement over the three factor structure, however still not fitting their data sufficiently well.

Overall, it seems that while the total score of the OQ may serve as an indicative reference for clinicians to identify clients at risk of deterioration, the inconclusive results of the OQ factor analyses call therapists to be cautionary in using the subscales as benchmarks for clinical decisions. This probably explains why although the OQ-Analyst software reports convey the OQ subscale values, its graphical representation is based solely on the OQ total score.

Method

Translation

First, the Outcome Questionnaire was independently translated from English into Polish by two translators. Discrepancies between these translations were resolved by discussion. Then, the agreed upon version was tested for clarity and adequacy on a group of eight certified psychotherapists and their patients. Next, in an effort to maintain semantic equivalence the back-translation, independent of the initial translation, was implemented by one translator. The accuracy of the back-translation was confirmed by the authors of the original measure. Finally, the questionnaire was again tested for comprehensibility and reliability on another group of seven Polish clinicians and their psychotherapy patients. All three translators were experienced clinicians and psychotherapy researchers. Their average age was 46.33 years (SD = 13.58, range 38–62), and their average years of experience was 20.33 years post licensure (SD = 8.26, range 14–32). All of them had PhDs, two of them were licensed psychologists and one was a psychiatrist. All three provided psychotherapy in both in – and outpatient settings. One of them described her treatment orientation as cognitive-behavioural, one as psychodynamic and another as integrative.

Participants

The study sample included three clinical subsamples and one non-clinical subsample (see Table 1). A clinical sample consisted of 211 day-patients coming from a particular psychotherapy centre from the south of Poland, 234 inpatients of a specific hospital-based psychotherapy clinic located in central Poland, and 137 psychotherapy outpatients from ten private practices from all over the country. All these subsamples were of convenience. For exclusion criteria (serious suicidal ideation, psychosis, major depression, active substance abuse) some respondents (37/619; 5.98%) were removed from the study, because their treatment was heavily based on pharmacotherapy and psychotherapy played a marginal role. All participants completed their first OQ as a part of the intake procedure prior to the first therapy session. Subsequent questionnaires were taken fifteen minutes before each consecutive therapy session for the whole duration of a particular treatment. In all the study subsamples, the questionnaires were distributed within a one week interval. An average number of 10.45 (SD = 4.19) questionnaires were completed by day-patients, 11.31 (SD = 2.04) by inpatients, and 13.84 (SD = 7.16) by outpatients.

Eighteen licensed psychologists, nine psychiatrists, and ten psychotherapists in training provided treatments during the data collection. The average age of the therapists, in all three settings, was 39.35 years (SD = 7.59, range 28–52), and the average

years of experience post licensure was 11.37 (SD = 8.5, range = 3–26). Patients were assigned to therapists using therapist availability and clinical factors. Most (21/37, 56.76%) of the participating therapists described their treatment orientation as integrative, while the remaining clinicians adhered to psychodynamic (eight), cognitive behavioural (six) or systemic (two) tenets. Adherence checks of treatment integrity were not performed. Therapists were not advocates of the questionnaire used, and were not reimbursed for its use.

The non-clinical subsample was based on 426 students from seven Polish universities. Importantly, they were not students of psychology, pedagogy, social sciences, or medicine. These respondents were recognised as non-patients due to the fact that at the time of the data collection they were not participating in any form of psycho – or pharmacotherapy. The questionnaires were distributed by teaching assistants on a weekly basis for four subsequent weeks. Respondents (57/483; 11.8%) who completed the OQ less than three times were not included in the data analysis. Participation was voluntary and anonymous. An average number of 3.98 (SD = 0.1) questionnaires were completed by the non-patients. Neither extra credit nor any type of reimbursement were offered to non-patients.

Table 1. **Characteristics of the Sample**

Sample		Gender		Age	
	n	Female n (%)	Male n (%)	Range	Mean (SD)
Day-patients	211	143 (67.8)	68 (32.2)	19–57	31 (8.49)
Inpatients	234	170 (72.6)	64 (27.4)	18–57	31 (8.71)
Outpatients	137	86 (62.8)	51 (37.2)	20–61	32 (7.82)
Patients – total	582	399 (68.6)	183 (31.4)	18–61	31 (8.44)
Non-patients	426	231 (54.2)	195 (45.8)	18–59	23 (6.79)
Whole sample (Patients and Non-patients)	1008	630 (62.5)	378 (37.5)	18–61	28 (8.66)

SD – Standard Deviation

Data Analysis

Initially, the reliability of the original three factor model on Polish data was tested with the assumption of the orthogonal model. However, since the explained variance and CFA indexes indicated a poor fit of the original three factor model to the Polish data, we decided to search for other alternative structures of the OQ-45.

Data analysis consisted of the following major steps: 1) parallel analysis (PA) to determine the number of factors identified with the OQ subscales; 2) exploratory

factor analysis (EFA) to define the items forming each OQ subscale; 3) confirmatory factor analysis (CFA) to confirm the OQ structure as suggested by the PA and the EFA; 4) correlation analysis to examine the interrelationships between the OQ subscales; 5) criterion equivalence to distinguish the clinical from the non-clinical sample; 6) clinical significance to establish the cut-off scores and reliable change indexes; and 7) test-retest reliability to measure the OQ sensitivity to change.

Parallel Analysis as simulation based on the Monte Carlo method compares the observed eigenvalues with those obtained from uncorrelated variables. A particular factor is retained if its associated eigenvalue is larger than the 95th percentile of the eigenvalues distribution derived from the random data. Also, the value of the first eigenvalue being larger than the second eigenvalue indicates the number of factors to be retained [26].

Factorial validity of the OQ-45 was used for two purposes: to confirm the number of retained factors and to find items corresponding to appropriate factors. Factorial validity was examined with the EFA of varimax, promax, and oblimin rotations, conducted per person on the average between the first and the last administration; next with the CFA of covariance matrices based on maximum likelihood (ML) or the generalised least square (GLS) method of estimations relied on the data derived from the fourth session (due to the number of non-patients sessions). We decided to apply both ML and GLS in order to allow comparability of our results with a wider range of authors, who relied on either ML or on GLS .The ML method was used e.g. by Bludworth et al. [18], while e.g. de Jong et al. [9] and Mueller et al. [21] used the GLS method. Due to the lack of a normal distribution of our subscales we applied ML with robust standard errors and robust test statistics. Additionally, we recognised ML as an acceptable method because of the large number of records.

The following goodness-of-fit indexes were applied in the CFA: the chi-square (χ^2), the chi-square divided by degrees of freedom (χ^2/df), the root mean square residual (RMR), the root mean square error of approximation (RMSEA), the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the normed fit index (NFI), the comparative fit index (CFI), the Tucker-Lewis index (TLI), the critical N (CN), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). We decided to report all of these indexes for the sake of comparability with other studies. We followed the generally recognised guidelines indicating a close fit: a χ^2 being non-significant; χ^2/df < 3; the RMR < 0.05; the RMSEA < 0.07; the GFI, the AGFI, the CFI, the NFI, and the TLI > 0.9; the CN > 200; while the AIC and the BIC indicating with its lowest value the most adequate fit and parsimony of the model [27].

In the exploratory factor analysis (EFA) the number of factors to be extracted was determined by using a combination of the number of factors (based on PA), polychoric correlation matrix, the eigenvalues (greater than 1), the scree plot, psychological theo-

ries, as well as clinical arguments. Items with substantial factor loadings in absolute value of 0.30 or larger were retained in the original subscales. Consequently, all three original subscales were shortened and some items were re-located from one original subscale to another. New subscales were formed out of the remaining items. The CFA goodness-of-fit indexes confirmed the adequacy of these modifications.

In the next step, we analysed, as suggested by literature [28], the bi-level model, considering several different bi-level structures of the OQ-45. Then, these structures were content analysed by the Polish experts in the field. These procedures left us with two separate bi-level models, which seemed to be the most adequate fits for the Polish data set: the bi-level model of the total score and five orthogonal subscales, called later "five factor bi-level model"; and the bi-level model of the total score, five orthogonal subscales, and two oblique factors, called later "seven factor bi-level model". Since goodness-of-fit indexes favoured the seven factor model over the five factor structure, we decided to use the former for further analyses.

Correlation coefficients between the subscales were computed with the assumption that the values greater than 0.31 and lower than 0.6 indicate a moderate relationship between the subscales, while a correlation above 0.6 for a subscale and the total score would demonstrate their high cohesion. Criterion equivalence was estimated as differences in scoring for the clinical and non-clinical samples. Wilks' λ test was applied. Crucial values for estimating clinical significance, cut-off scores and the Reliable Changes Indexes, were established with the usage of Jacobson and Truax [29] criteria modified by Lambert, Hansen and Bauer [30]. Reliability was estimated in terms of internal consistency (Cronbach's alpha) and test-retest (Pearson product-moment correlation coefficient). Coefficients above 0.7 have been recognised as sufficient values of reliability. The time difference used to conduct the test-retest stability was one week. Sensitivity to change was measured with the usage of paired t-tests (Student's t-tests) and the effect size statistics. The standardised *d* effect size was applied. Effect sizes were weighted by sample sizes. Classification of d (large > 0.79, medium 0.50–0.79, or small < 0.50) was used [31].

The EFA was conducted in SPSS 21, STATISTICA 9.0, and R 11.0; the CFA in AMOS 21, and R 11.0; test-retest and internal consistency in STATISTICA 9.0, and SPSS 21. Missing data were replaced with values according to the Monte Carlo procedure [32]. Less than one percent of missing values per question was found (0.93%).

Results

Parallel analysis

The PA results pointed out to the most accurate structure of the Polish OQ, i.e., consisting of seven subscales. This was the case at every data record (i.e., each ses-

sion), for all the three groups of patients, treated as independent as well as combined subsamples. Factors corresponding to the eigenvalues greater than 1 were retained. For the lack of space only the results for the whole study group are presented (see Figure 1). First seven adjusted eigenvalues are depicted with solid line, while seven random eigenvalues with dotted line.

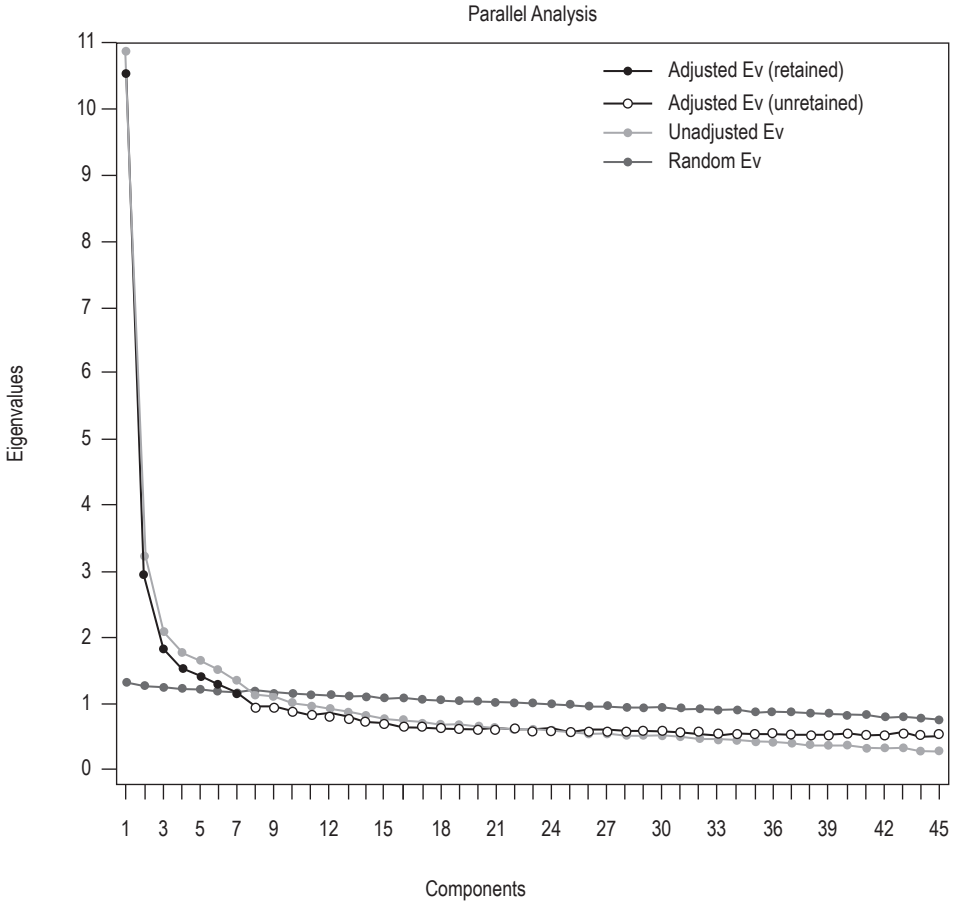

Figure 1. **Parallel analysis: Eigenvalues from actual and random data for retained factors (N = 1008).**

Factor analysis

Inconclusive results of previous studies as well as cultural divergences between American and Polish populations [33], made us assume that the factor structure of the Polish OQ may have been different than the original American model, hence we decided to test it with the exploratory factor analysis (EFA). In confirmatory factor analysis (CFA), we were following suggestions to rely both on empirical (i.e., modification of item structure) and rational (i.e., relevance to psychological theories and adherence to clinical practice) procedures in an effort to obtain more adequately fitting models [21].

The three factor model, as proposed by the OQ developers, was not positively verified with the EFA, and there was only a partial match of the item distribution between American and Polish versions. Furthermore, it poorly fulfilled goodness-of-fit criteria of the CFA. Factor loadings of the Polish data indicated that while some of the items of the three factor model still belonged to original first order orthogonal subscales (Table 2), others were re-located from one original subscale to another, while remaining items were found more suitable for a second order orthogonal factors. Consequently, original subscales were modified, yet since they still seemed to measure the originally adhered concepts, they kept their names.

Table 2. **Standardised Factor Loadings of the Seven Factor Bi-level Model (N = 1008)**

Item	IR	SD	SR	SC	AB	SA	SR2	TOT
OQ1[a]	0.62					0.44[x]		0.56
OQ7[a]	0.36							0.76
OQ12[c]	0.48		0.39[x]				0.4	0.6
OQ13[b]	0.88							0.22
OQ15[b]	0.65	0.33[x]					0.54	0.46
OQ17[a]	0.42							0.77
OQ18[a]	0.64					0.48[x]		0.52
OQ20[a]	0.83							0.29
OQ21[c]	0.72						0.55	0.45
OQ24[b]	0.75					0.57[x]		0.43
OQ30[a]	0.5			0.43[x]		0.41[x]		0.49
OQ31[b]	0.86							0.25
OQ37[a]	0.76							0.37
OQ43[a]	0.79							0.36
OQ2[b]		0.52	0.36[x]			0.42		0.58
OQ5[b]	0.44[x]	0.47				0.42[x]		0.58

table continued on the next page

OQ6[b]	0.29[x]	0.3		0.35[x]		0.31[x]		0.69
OQ8[b]	0.29[x]	0.29						0.74
OQ9[b]	0.39[x]	0.46	0.32[x]			0.49		0.51
OQ10[b]		0.6				0.47		0.53
OQ16[a]		0.45					0.3	0.7
OQ23[b]		0.33						0.56
OQ25[b]		0.48				0.35		0.65
OQ27[b]		0.69				0.51		0.49
OQ29[b]		0.67				0.46		0.54
OQ33[b]		0.66				0.46		0.54
OQ34[b]		0.69				0.52		0.48
OQ35[b]		0.42				0.2		0.8
OQ36[b]		0.58				0.44		0.56
OQ40[b]	0.34[x]	0.44					0.37[x]	0.63
OQ41[b]		0.62				0.44		0.56
OQ42[b]		0.44						0.51
OQ45[b]		0.7				0.52		0.48
OQ3[b]	0.38[x]		0.4				0.37	0.63
OQ4[c]			0.51				0.39	0.61
OQ22[b]		0.36[x]	0.58			0.47[x]		0.53
OQ28[c]			0.67				0.48	0.52
OQ38[c]			0.71					0.44
OQ14[c]				0.35			0.21	0.79
OQ19[a]				0.55			0.32	0.68
OQ39[c]			0.44[x]	0.58				0.44
OQ44[c]				0.57		0.38[x]		0.62
OQ11[b]					0.56		0.35[x]	0.65
OQ26[a]					0.65		0.43[x]	0.57
OQ32[c]					0.6		0.37[x]	0.63

Bartlett's Test of Sphericity: $\chi^2 = 23\ 887$; df = 990; $p < 0.0001$; KMO Measure: 0.967.

IR – Interpersonal Relations subscale; SD – Symptom Distress subscale; SR – Social Role Performance subscale; SC – Social Conflicts subscale; AB – Addictive Behaviours subscale; SA – Somatisation and Anxiety subscale; SR2 – Social Role 2 subscale; TOT – total score; a – originally part of the IR subscale of the American OQ-45.2; b – originally part of the SD subscale of the American OQ-45.2; c – originally part of the SR subscale of the American OQ-45.2; x – standardised factor loadings greater than 0.3, but smaller than the largest factor loadings for a given item

The Symptom Distress (SD) factor contains 18 out of 25 items of the original SD subscale, and one item originally placed in the IR subscale (Table 2). Such newly defined item structure of the SD factor closely resembles the Psychological Distress subscale of Minami et al. [20]. This subscale, however, did not contain the OQ16 ("I am concerned about family troubles"), which was found to be part of Polish SD subscale. In our opinion, this item is more of an adequate component of the SD subscale than the IR, since it seems to adhere rather to a cognitive and emotional phenomenon of concern, which potentially can rise to the level of symptomatic distress; than to measure troubling dynamic within the family system. Besides, since Poles, on average, more than Americans, tend to express their discontent and to dwell deeper in their concerns [33], they may be more prone to address the OQ16 rather from symptomatic than the interpersonal perspective. Thus we recognised the OQ16 as adequate component of the factor that evaluates various features of psychological and physical symptoms.

Fourteen items loaded substantially on the Interpersonal Relations (IR) factor, which measures different dimensions of the interpersonal functioning. Eight items of the Polish IR are equivalent with the American IR, two items are originally located in the SR subscale while four in items originally constitute the SD (Table 2). The presence of two former SR items (OQ12, "I find my work/school satisfying"; OQ21, "I enjoy my spare time"), in the modified IR subscale is explainable, if one takes into account that the extent to which people find their interpersonal relations satisfying often is interlocked with the way people experience their professional and leisure activities [34].

Analogically, items originally placed in the SD subscale (e.g., OQ13, "I am a happy person"; OQ15, "I feel worthless") not only seem to directly relate to the level of symptoms, but also to the fact that the quality of interpersonal functioning depends on a person's self-esteem and general satisfaction with life [35]. Importantly, the item structure of the Polish version of the IR very closely resembles the Negative Self-Worth subscale of Kim and colleagues [19].

Another factor is built of five items that relate to functioning in a social role. Thus this factor is named, identically to the American OQ, the Social Role (SR) factor. It holds three original SR items (Table 2) and two items from the original SD subscale (OQ3, "I feel no interest in things"; OQ22 "I have difficulty concentrating"). Since, these two latter items refer to obstacles which usually significantly hamper social functioning [36], we found them adequate for the SR. Noteworthy, the item structure of the Polish SR subscale closely reflects the structure of the Loss of Interest factor of Kim et al. [19].

The remaining items constitute two orthogonal factors, which were not conceptualised by the OQ developers. One of these factors is named the Addictive Behaviours (AB) factor since it seems to adequately capture characteristics of an addicted indi-

vidual. It contains three items: OQ11, "After heavy drinking, I need a drink the next morning to get going", originally in the SD factor; OQ26, "I feel annoyed by people who criticise my drinking or drug use", originally in the IR factor; and OQ32 "I have trouble at work/school because of drinking or drug use", originally in the SR factor (Table 2). The fact that the American version of the questionnaire does not contain a separate AB subscale might have occurred due to the fact that the OQ was normalised in Utah, the state which is recognised as the most sober in the US. Noteworthy, our AB subscale perfectly matches, item content wise, substance abuse related factors extrapolated by other researchers [19, 20]. It could be therefore hypothesised that the AB factor is not specific just to Polish society, but also to these populations where drinking is more socially accepted.

The other new orthogonal factor is loaded by four items that measure conflicts in various social situations, thus it is named the Social Conflict (SC) factor. Three of its items were conceptualised in the original version of the SR subscale: OQ14, "I work/study too much": OQ39 "I have too many disagreements at work/school": and OQ44 "I feel angry enough at work/school to do something I might regret": while one item was originally placed in the IR subscale: OQ19 "I have frequent arguments" (Table 2). The item structure of the SC greatly resembles the Interpersonal Conflict factor as defined by Minami et al. [20]. Noteworthy, this subscale does not hold the OQ14. It is debatable to what extent the presence of this particular item in the SC is adequate, not only due to its relatively low factor loading, but also content wise. For the same reasons, OQ14 frequently was found problematic also in other studies, to the extent that its content modification or even removal from the OQ has been suggested [9, 12, 19–21]. Nevertheless, a separate subscale measuring social conflicts, independently of the SR and IR factors could be of clinical relevance especially for societies like the Polish, which allow for open expression of frustration and encourage engagement in social conflicts [33].

Having observed that certain items could be adequately placed in more than one factor, we decided to test whether a model with oblique factors would be a more adequate fit to our data. Thus, we computed correlation matrix. Its high values, all above 0.7, suggested the presence of additional non-orthogonal factors. Then, each item was indicated to one of the already isolated factors (IR, SD, SR, SC, AB). Next, with the usage of oblique rotation based on promax and direct oblimin method, each item was separately related to potential oblique factors. Once, the new oblique factors were statistically formed, their contents were analysed both from the theoretical and clinical perspective.

The first oblique factor is labelled Somatisation and Anxiety (SA), since it is loaded with twelve items measuring symptoms specific for anxiety and somatic disorders, e.g., OQ10, "I feel fearful" or OQ41, "I have trouble falling asleep or staying asleep" (Table 2). Although all the SA items are also to be found in the SD subscale,

we recognise its isolation as clinically relevant especially for these respondents who are diagnosed with neurotic and anxiety disorders. Since, they form the majority of psychotherapy patients, at least in outpatient settings, we assert that the SA subscale could be a useful addition to the five factor structure. The fact that the SA factor mirrors the Anxiety and Somatic Distress factor of de Jong et al. [9] and closely resembles the Anxiety-Somatisation factor of Kim et al. [19] suggests that the SA could be specific not just for Polish respondents.

The other oblique factor is named the Social Role 2 (SR2), since it is created out of nine items, which measure some specific aspects of the social role performance, i.e., items located also in the SR factor e.g., OQ28, "I am not working/studying as well as I used to"; items placed also in the IR e.g., OQ21, "I enjoy my spare time"; items coming from the SC factor e.g., OQ14, and item common with the SD subscale OQ16 (Table 2). Although most of its factor loadings are of sufficient value, its clinical applicability seems to be limited to those patients who are entangled in their social functioning.

In the next step the CFA was conducted in order to validate: a) the original three factor orthogonal model; b) the five factor bi-level model; and c) the newly identified seven factor bi-level model (Table 3). The three factor model tested on the Polish data poorly met goodness-of-fit criteria.

Even for the more favourable GLS-based method only two out of nine indexes (RMSEA, CN) met the CFA recommended criteria at a satisfactory level. The relatively high values of the AIC and the BIC suggested inadequate fit and lack of parsimony for that model. The five factor bi-level model of a total score and five orthogonal subscales when explored with the GLS-based CFA was supported by three indexes meeting criteria at a satisfactory level (χ^2/df, RMSEA, CN) and one index (GFI) approaching criteria at an acceptable level. The AIC and the BIC suggested more favourable fit for this solution than for the three factor structure.

The bi-level structure of a total score, five orthogonal and two oblique subscales, when tested with the GLS-based CFA, seems to be the best fitting model to our Polish data (Figure 2). Three goodness-of-fit indexes (χ^2/df, RMSEA, CN) fulfilled the recommended criteria. Two indexes (RMR, GFI) approached the CFA criteria at an acceptable level. Moreover, this model, of all tested structures, was indicated as the most adequate fit to our data, also due to the lowest value of the AIC and the BIC. Overall, the fit of the seven factor model, according to both GLS and ML methods, was much improved over the five factor and over the original three factor models. The seven factor bi-level model indicates the best fit to our data based on the proportion of variance. Cumulative variance (varimax rotation) of the orthogonal three factor solution yields the value of 35.44%. For the four factor orthogonal solution the value of variance is 47 %. More variance is explained with the orthogonal five factor model, i.e. 50.01 %, and even more with seven factor bi-level structure (55.04%).

Table 3. **Confirmatory Factor Analysis Goodness-of-fit Indexes (N = 1008)**

			Indexes											
Model	Estimation	χ^2	df	χ^2/df	RMR	RMSEA	GFI	AGFI	NFI	CFI	TLI	CN	AIC	BIC
3-factor original model	GLS	3125[a]	942	3.32	0.087	0.049*	0.856	0.842	0.15	0.187	0.146	314*	3311	3764
3-factor original model	ML robust	6210[a]	942	6.59	0.064	0.076	0.712	0.683	0.747	0.776	0.765	158	6396	6849
5 factor bi-level model (with total score)	GLS	2670[a]	890	2.66*	0.056	0.041*	0.891**	0.873	0.356	0.449	0.361	392*	2659	3366
5 factor bi-level model (with total score)	ML robust	3479[a]	890	3.76	0.041*	0.053*	0.856	0.833	0.863	0.896**	0.884	278*	3638	4344
7 factor bi-level model with five orthogonal (with total score) and two oblique factors	GLS	2257[a]	865	2.6*	0.054**	0.041*	0.9**	0.877	0.386	0.482	0.407	280*	2597[b]	3326[b]
7 factor bi-level model with five orthogonal (with total score) and two oblique factors	ML robust	3082[a]	865	3.56	0.038*	0.05*	0.869	0.843	0.874	0.91*	0.883	294*	3418	4237

* Meets the recommended criteria (satisfactory level).
** Approaches the recommended criteria (acceptable level)
[a] $p < 0.001$; [b] indication of the most adequate fit and parsimony of all tested models.

GLS – the generalised least square; ML – maximum likelihood; χ^2 – chi-square; df – degrees of freedom; RMR – root mean square residual; RMSEA – root mean square error of approximation; GFI – goodness-of-fit index; AGFI – adjusted goodness-of-fit; NFI – normed fit index; CFI – comparative fit index; TLI – the Tucker-Lewis index; CN – critical N; AIC– the Akaike information criterion; BIC – the Bayesian information criterion.

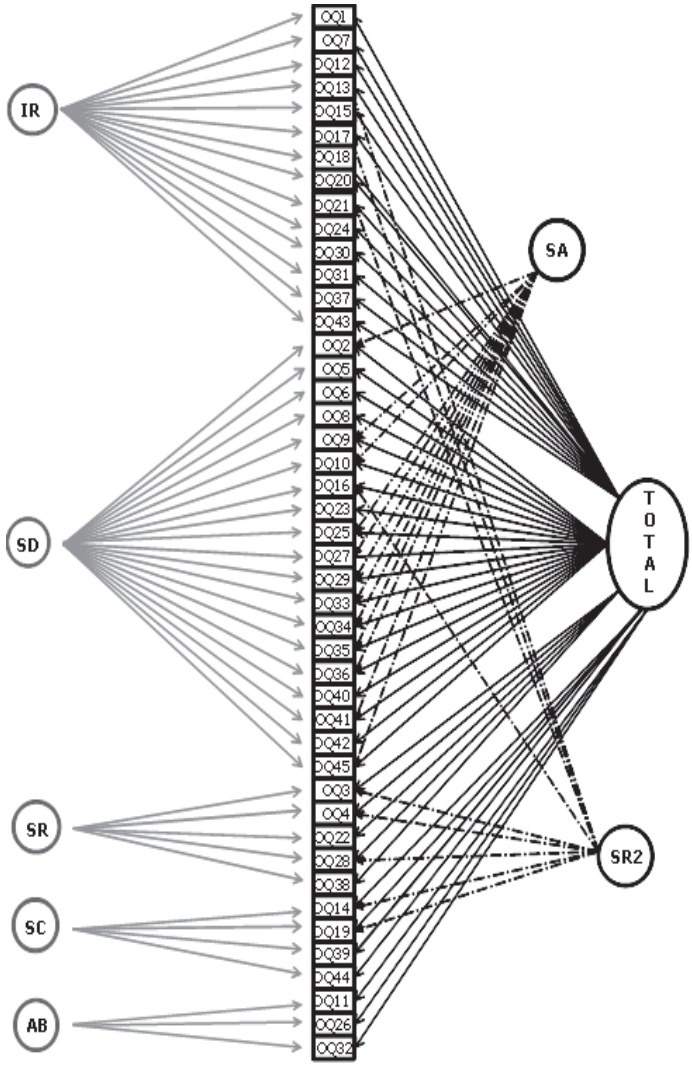

Figure 2. **Seven Factor Bi-level Model of Total Score, Five Orthogonal Factors and Two Oblique Factors**
IR – Interpersonal Relations; SD – Symptom Distress; SR – Social Role Performance; SC – Social Conflicts; AB – Addictive Behaviours; SA – Somatisation and Anxiety; SR2 – Social Role 2; N = 1008

Correlations

Correlations between the subscales reflect multidimensionality and independence in the factor structure, indicating overall, moderate construct validity. On one hand, two

orthogonal subscales of the second order (SC, AB), consistently obtain values below 0.6 (Table 4). These findings seem to indicate the cohesiveness of these two factors. All the other factors however correlate above 0.6. It is not surprising in the case of the oblique factors (SA, SR2), which by definition cover constructs already tackled by other factors (check for especially high correlation between SA and SD). These correlation values support the oblique structure of the Polish version of the OQ-45. High correlation values for the first order orthogonal factors (SD, IR, SR) are of concern though. These results may indicate that these subscales, even being modified, may be measuring overlapping constructs. Correlations between subscales and the total score are usually above 0.80, which speaks to cohesiveness of the measure. The exception of two oblique factors most likely occurred due to the small number of items in the SC and the AB subscales.

Table 4. **Correlations between the Subscales and the Total Score**

N = 1008	TOT	IR	SD	SR	SC	AB	SA	SR2
TOT	1							
IR	0.91	1						
SD	0.95	0.78	1					
SR	0.82	0.68	0.74	1				
SC	0.47	0.33	0.41	0.36	1			
AB	0.12	0.06	0.07	0.08	0.15	1		
SA	0.91	0.71	0.98	0.7	0.39	0.07	1	
SR2	0.9	0.81	0.82	0.85	0.51	0.07	0.76	1

IR – Interpersonal Relations subscale; SD – Symptom Distress subscale; SR – Social Role Performance subscale; SC – Social Conflicts subscale; AB – Addictive Behaviours subscale; SA – Somatisation and Anxiety subscale; SR2 – Social Role 2 subscale; TOT – total score

Criterion of equivalence

Table 5 shows differences in scoring (between the first and the last session) of the Polish clinical and non-clinical samples.

Table 5. **Means and Standard Deviations of OQ in the Polish Clinical and Non-clinical Samples**

	Clinical group		Non-clinical group (university)	
	n = 582		n = 426	
	Mean	St.D	Mean	St.D

table continued on the next page

	Clinical group		Non-clinical group (university)	
	n = 582		n = 426	
IR	26.28	7.59	13.91	6.88
SD	33.8	9.99	20.51	8.86
SR	9.87	3.31	6.79	2.85
SC	4.82	2.18	4.49	1.88
AB	0.47	1.02	0.59	1.17
SA	20	6.8	12	5.6
SR2	17	4.3	12	4.1
TOTAL	75.22	19.39	46.28	18.23

IR – Interpersonal Relations subscale; SD – Symptom Distress subscale; SR – Social Role Performance subscale; SC – Social Conflicts subscale; AB – Addictive Behaviours subscale; SA – Somatisation and Anxiety subscale; SR2 – Social Role 2 subscale; TOTAL – total score. St.D – Standard Deviation

The sample of Polish non-patients when compared to the sample of Polish patients has significantly better level of functioning on all subscales and the total scale, i.e. Wilks' $\lambda = 0.52$, F (7, 1000) = 133.76, $p < 0.00001$. Effect sizes for the difference between the clinical and community samples are usually very large (e.g., for the IR F (1, 1006) = 745.14, $p < 0.00001$, d = 1.7, 95% CI = [1.25, 2.15] and for the total scale F (1, 1006) = 246.21, p <0,00001, d = 1.53, 95% CI = [0.37, 2.7]), with an exception of the SC and the AB, which yield small and nonsignificant values: for the SC F (1, 1006) = 8.47 $p < 0.0037$, d = 0.16, 95% CI = [0.03, 0.29] and for the AB F (1, 1006) = 2.93, $p < 0.0874$, d = – 0.11, 95% CI = [-0.18, – 0.04].

The mean scores for the Polish non-clinical sample are above the American equivalents when compared to the US university sample (t = 12.4, $p < 0.000001$, d = 0.68, 95% CI = [-1.13, 1.59]) or the US community sample ($t = 9.4$, $p < 0.000001$, $d = 0.48$, 95% CI = [-0.98, 1.1]). The effect size for comparison between the Polish clinical sample and the US clinical sample is very small and statistically nonsignificant (t = 1.86, p = 0.18, d = 0.05, 95% CI = [-1.7, 0.9]).

Clinical significance and reliable change

For the Polish OQ the cut-off scores are the following: 20 for the IR subscale, 27 for the SD, 8 for the SR, 5 for the SC, 1 for the AB, 16 for the SA, and 14 for the SR2, and for the total score 60. Scores above the cut-off point indicate that an individual belongs to the clinically dysfunctional range. In other words, the cut-off score of 60

indicates the sensitivity for the OQ total score at the level of 0.90, which means that 90% of the non-patient sample belongs to the functional sample while 90% of clinical sample is adequately identified as dysfunctional. The Reliable Change Indexes (RCI) are the following: 8.27 for the SD, 5.73 for the IR, 4.46 for the SR, 4.64 for the SC, 2.45 for the AB, 7.31 for the SA, and 6.72 the SR2. The RCI for the OQ total score is 12.9. In other words, a patient has to improve a minimum of 13 points on the OQ total scale to obtain a reliable change.

Internal consistency and test-retest

Internal consistency and test-retest reliability of the Polish version of the OQ are of sufficient values for subscales and the total scale, indicating adequacy of the seven factor bi-level model. Values of internal consistency (Table 6) in all of the cases except for the SC are at the level of 0.7 or above. Most test-retest reliability values (Table 7) are above 0.7 and some are above 0.8. The reaming ones yield values above 0.6, which is recognised as the minimal level of reliability of the instrument. Test-retest results point to stability in time of the Polish version of the OQ-45.

Table 6. **Internal Consistency: Cronbach's alpha**

Internal consistency	Cronbach's alpha					
	Clinical sample n = 582		Non-clinical sample n = 426		Total sample n = 1008	
Subscales	n	a	n	a	n	a
IR	582	0.88	426	0.91	1008	0.93
SD	582	0.91	426	0.92	1008	0.93
SR	582	0.80	426	0.78	1008	0.82
SC	582	0.61	426	0.60	1008	0.61
AB	582	0.71	426	0.70	1008	0.71
SR2	582	0.70	426	0.75	1008	0.76
SA	582	0.84	426	0.85	1008	0.88
TOTAL	582	0.93	426	0.95	1008	0.95

IR – Interpersonal Relations subscale; SD – Symptom Distress subscale; SR – Social Role Performance subscale; SC – Social Conflicts subscale; AB – Addictive Behaviours subscale; SA – Somatisation and Anxiety subscale; SR2 – Social Role 2 subscale; TOTAL – total score

Table 7. **Test-retest Reliability (Pearson Product-moment Correlation Coefficient)**

	Test-retest reliability					
	Clinical sample n = 582		Non-clinical sample n = 426		Total sample n = 1008	
Subscales	n	r	n	r	n	r
IR	582	0.76	426	0.84	1008	0.89
SD	582	0.77	426	0.84	1008	0.87
SR	582	0.7	426	0.72	1008	0.79
SC	582	0.63	426	0.68	1008	0.65
AB	582	0.73	426	0.69	1008	0.72
SR2	582	0.67	426	0.75	1008	0.8
SA	582	0.78	426	0.8	1008	0.86
TOTAL	582	0.77	426	0.87	1008	0.89

IR – Interpersonal Relations subscale; SD – Symptom Distress subscale; SR – Social Role Performance subscale; SC – Social Conflicts subscale; AB – Addictive Behaviours subscale; SA – Somatisation and Anxiety subscale; SR2 – Social Role 2 subscale; TOTAL – total score

Sensitivity to change

Sensitivity to change was measured between the first and fourth session, for the group of randomly selected 161 patients coming from all clinical subsamples. The Polish version of the OQ showed high sensitivity to change on all subscales: e.g., SD: $t(161) = 8.43$, $p < 0.001$, $d = 0.56$; IR: $t(161) = 2.04$, $p < 0.001$, $d = 0.15$; $d = 0.27$; AB $t(161) = 4.03$, $p < 0.001$, $d = 0.33$; SA: $t(161) = 8.02$, $p < 0.001$, $d = 0.53$; and the total scale $t(161) = 7.33$, $p < 0.001$, $d = 0.50$.

Discussion

Our study was the first attempt to establish the psychometric properties and factorial validity of the Polish OQ-45, as well as its equivalence with the original version of the measure. We found a poor fit of the original three factor OQ to our data, a reasonable fit of the five factor model, and a good fit of the seven factor model. The latter structure is a bi-level model of the total score, five orthogonal factors (subscales specific for the Polish OQ, i.e. Social Conflicts and Addictive Behaviours; and original yet modified subscales, i.e. Symptom Distress, Interpersonal Relations, and Social Role) and two oblique factors (Somatisation and Anxiety, Social Role

2). When compared to other evaluated solutions, not only does this model explain the largest percentage of the variance, its goodness-of-fit indexes meet the CFA recommended criteria most frequently. Its psychometric properties and sensitivity to change seem to be sufficient, its cut-off scores and clinically significant change indexes are adequate and similar to original measure. Overall, although the Polish OQ model does differ from the original structure, its equivalence with the American OQ could be recognised as satisfactory.

Fit of the original three factor model to Polish data

Noteworthy, the fit of the original three factor model to our data, when compared, within ML or GLS methods, to other validations was only slightly worse [9, 11] or even quite equivalent [21]. Still, we decided not to recognise the three factor model as an adequate fit to our data for the following reasons: (1) factor loadings based on the PA and EFA, did not support the original distribution of items into three subscales, but to the contrary, they formed new subscales, specific for Polish respondents; (2) the three factor model obtained substantially lower CFA indexes when compared to seven factor structure; and (3) the new subscales to a certain extent reflect item distribution and factor structure as obtained in other studies [9, 19, 20].

The fact that the original three factor structure when tested on our data was not positively verified with the EFA and only poorly fulfilled the CFA goodness-of-fit criteria is not surprising in the light of inconclusive results of previous studies dealing with the OQ factor structure [9, 11, 12, 18–21]. Their authors, as much as the OQ-45 developers [22], on one hand appreciate the OQ as a measure containing one general factor and multiple subscales, yet are aware that the subscales may not be that discrete as suggested by psychometric properties, they suggest other than three factor structures of the OQ, which potentially may more adequately capture the multidimensionality of individual functioning. It is therefore to be emphasised that the EFA findings and the CFA results along with correlations support the seven factor bi-level structure of the OQ-45 as most appropriate for the Polish data set.

Psychometric properties of the Polish version of the OQ

The Polish OQ is equipped with adequate cut-off scores and RCI useful for estimating clinical significance, and thus creates valid categories of patients who are recovered, improved, unchanged, or deteriorated during a treatment. The cut-off point of the Polish OQ is at the level of 60, i.e., the Polish respondents who get 60 points or less are therefore to be classified in normal range, while respondents with 61 points or more in dysfunctional range. The cut-off point of the Polish OQ is within the range of cut-off points of other OQ versions (Dutch 55, American 63, Spanish 66, Italian 66). The cut-offs for three original yet modified subscales in the Polish

OQ correspond to American ones (IR Polish 20, US 15; SD Polish 27, US 36; and SR Polish 8, US 12). The RCI for the OQ total score for the Polish OQ is 13, while for both American and Dutch is 14 [9, 22].

Three original subscales (SD, IR, SR) although shortened and modified, obtain psychometric properties comparable to the US and other versions of the OQ. Internal consistency of the Polish OQ is overall higher, with the exception of the SD which usually yields the same values, when compared to the US [22], Dutch [9], Italian [11], and Swedish [12] data. Since the majority of coefficients cross the threshold of 0.70, we claim that the Polish OQ is characterised with good internal consistency. The SR subscale, the only one holding values below 0.70, still has substantially higher internal consistency when compared to other language versions, including the American one. Test-retest reliability of the total score as well as three first order orthogonal subscales of the Polish OQ, when compared to the US and Dutch data, usually is higher, for both clinical and non-clinical samples. Its satisfactory values attest to the Polish OQ's stability over time in scoring.

The IR and the SR consistently correlate higher and the SD slightly lower with the total score, when compared to Dutch [9], German [10], and Italian [11] versions of the OQ. Unfortunately, the OQ manual [22] does not provide correlations between the subscales and the total score for the original version. Overall, it could be claimed that these correlations support our decision to modify and shorten the original subscales, for they indicate high cohesion of the subscales with the total scale. The highest correlation of the SD with the total score, both in our study as well as in other validations [9, 10, 21], could indicate that the OQ is more a measure of symptom distress, rather than of interpersonal functioning or social role.

The CFA goodness-of-fit indexes indicate good fit of the seven factor bi-level structure to Polish data. Additionally, this fit when compared to the fit of the three factor model in other studies consequently yields more favourable indexes also more frequently meeting the recommended criteria. For example the Polish OQ structure has better fitting values of GFI, AGFI, NFI, CN, and worse χ^2/df when compared to the original OQ [21]; better χ^2/df, RMR, RMSEA, GFI, and worse NFI, CFI than the Dutch version [9]; and more favourable RMR, RMSEA, and AGFI, but worse CFI and AIC than the Italian OQ [11]. Although our seven factor model, when compared to solutions from other publications seems to have quite adequate approximation to the available data, still three indexes did not approach recommended criteria (χ^2, NFI, CFI). Since the χ^2 is dependent on the sample size, thus taking into account the relatively large sample size of our study, a significant χ^2 is not necessarily a sign of a poor fit. The low results of the NFI, which is partially based on χ^2, may also be accounted for by the sample size. The low results of the CFI are usual for a measure consisting of multiple items [37].

Overall these results speak to certain equivalence between Polish model and other language versions, and consequently indicate both statistical and clinical relevance of the Polish set of subscales.

Subscales specific to Polish OQ

The Addictive Behaviours subscale with its cut-off of zero and RCI of three is of high clinical relevance, almost reaching level of diagnostic sensitivity. To be recognised as an individual having considerable problems with alcohol and/or drugs it is enough to obtain just one point on the AB subscale. To achieve reliable change one needs to change by all available three points. Of course, since the AB factor only consists of three items, its psychometric properties, RCI, cut-offs, and correlations could be to a certain extent skewed. Still, its factor loading, internal consistency, and test-retest reliability make it comparable with other subscales. Nevertheless, the AB seems to be a clinically justifiable split of the original three factor structure.

The clinical utility of the Social Conflict subscale is not that implicit for the AB, and could be attributed to its relative heterogeneity. Probably the factor measuring such multicontextual phenomenon as social conflicts could benefit from addition of new items and/or from content modification of already existing questions. Noteworthy, the psychometric properties of the SC are hampered by the problematic question of OQ14. This item seems to be the least homogeneous of all SC items. Its removal from the OQ would increase the Cronbach's alpha up to 0.68 for the total scale, to 0.69 for clinical, and to 0.63 for the non-clinical subsample. Noteworthy, the same effect, i.e. increase of the Cronbach's alpha after discarding the OQ14, has been also noted in other validations [12].

The relatively poor psychometric properties of the SC, the AB, as well as the SR2 could be attributed to a small number of items in each of these subscales. Their reliability and internal consistency were usually better when clinical and non-clinical subsamples were combined into the total sample.

The SA subscale proved to be of clinical and statistical relevance. Its homogenous items seem to adequately measure phenomena related to somatisation and anxiety. Interestingly enough, although in terms of item distribution the SA subscale matches the Anxiety and Somatic Distress (ASD) factor of the Dutch OQ [9], it obtains slightly yet consistently better psychometric properties, when it comes to factor loadings, correlations with original subscales, internal consistency for the non-clinical population, and test-retest reliability. There is a perfect match between the SA and the ASD in terms of the RCI, correlation with total score, and internal consistency for clinical population.

The SR2 factor has, when compared to other subscales, more or less equivalent psychometric properties. Still, its relative heterogeneity questions its clinical utility and may call for future researchers to make modifications or remove it from the structure of the Polish OQ.

Strengths and limitations

We believe that this study provides results which are valuable not just to therapists who consider using the Polish version of the OQ-45 in their clinical practice, but also to researchers who develop psychotherapy outcome measures. The strengths of this study include: (1) a balanced and reasonably comprehensive account of previous studies and conflicting results found in analyses of the OQ; (2) a large and acceptably heterogeneous sample of clinical and non-clinical respondents; (3) clinical settings which resemble a treatment-as-usual environment; (4) analysis of data across multiple time points; (5) several carefully conducted analyses of possible factorial structures of the instrument; (6) the first known reliability estimates of the Polish version of the OQ; and (7) discussion elaborating upon how the role of cultural differences and the passage of time on the measure structure might influence the results. The generalisations from this study are limited in a number of ways: (1) not a very large size for each particular subsample; (2) the non-clinical subsample formed almost entirely by young adults; (3) lack of a normal distribution; (4) small effect sizes weakening the sensitivity of all the subscales; and (5) only 55.04% of the variance was explained with the seven factor bi-level structure.

Due to the large number of tests performed in this project the authors decided to explore effectiveness and stability of this tool in the other paper [38], which presents measures of the effectiveness based on Markov chains.

Conclusions

The paper presents results of Polish adaptation – based on factor analysis – of American Outcome Questionnaire (OQ-45.2). In summary, the seven factor bi-level model of the Polish OQ, seems to be – within the limitations of this study – both an empirically and clinically relevant measure of specific areas of client functioning (in terms of interpersonal relations; social role performance; social conflicts; symptom distress; somatisation and anxiety; and addictive behaviours). Psychometric properties of the Polish OQ are consistent with the American OQ and other language adaptations. Reliability and validity estimates of all subscales, including the new ones, are promising, and thus indicate our version of the OQ as sensitive enough of a tool for Polish respondents. Sensitivity to change of the Polish OQ is sufficient, allowing for discrimination between dysfunctional and non-clinical respondents. Specific cut-off scores allow for estimating reliable clinical change specific for Polish patients. Still, in its clinical application, as much as in the case of the original OQ, rather than on the subscales, it is safer to rely on the OQ total score. Overall, although the original subscales (SD, IR, SR) have been modified and shortened and additional factors had been modelled (AB, SC, SA, SR2), equivalence of the Polish OQ with the original version of the OQ is satisfactory.

Noteworthy, the clinical usage of the Polish version of the OQ 45-2 – analogically as in the case of the original measure – seems to be more reliable when relied upon just the total score rather than individual subscales. If only for the fact, that the American three factor structure – which served as a frame of reference in this study – is not a result of exploratory factor analysis, but rather arbitrary decisions of the authors of the OQ [1].

Differences between Polish and American factor structures could be attributed to numerous cultural divergences between Polish and American societies [33]. This assumption is somewhat backed up with differences in scoring of the Polish clinical and non-clinical subsamples, indicating that while patients from both countries are quite similar in their psychological characteristics as measured by the OQ, non-patients do differ at a significant level. The passage of time is another potential explanation of poor fit of the original three factor model to the Polish data. Since the questionnaire was designed nineteen years ago [1], it is likely that its structure may have been adequate then, but nowadays does not that precisely capture the experience of individuals. This could be especially relevant for Polish society, which has been substantially changing during the past twenty five years, being stretched between collectivistic and individualistic values, moving from survival to achievement oriented behaviours, dwelling more and more in a technology and Internet-based way of life [39]. It is likely that questions directly tackling such issues (e.g., work-home balance, on-line networks) could further increase relevance of the OQ not just for the Polish population.

Since this is a preliminary adaptation, there is a need for future studies further investigating validity of the factors found to be specific for the Polish OQ and testing its conceptual equivalence with the original OQ. It is to be noted that concurrent validity estimates for the Polish OQ with other measures remains to be tested. A larger number of respondents in each subsample may help to establish the cut-off points and the RCI for more homogeneous groups of respondents, i.e., inpatients, outpatients, or patients with specific diagnosis. Last but not least, we hope that our results will spark a discussion about whether or not efforts need to be taken in order to revise the content of the questions and factor structure of the original OQ.

References

1. Lambert MJ, Burlingame GM, Umphress V, Hansen NB, Vermeersch DA, Clouse GC. et al. *The reliability and validity of the Outcome Questionnaire*. Clin. Psychol. Psychother. 1996; 3: 249–258.

2. Hatfield DR, Ogles BM. *The use of outcome measures by psychologists in clinical practice*. Prof. Psychol. Res. Pract. 2004; 35: 485–491.

3. Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. *Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions*. J. Clin. Psychol. 2005; 61: 165–174.

4. Slade KL, Lambert MJ, Harmon SC, Smart DW, Bailey R. *Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools*. Clin. Psychol. Psychother. 2008; 15: 287–303.

5. Simon W, Lambert MJ, Harris MW, Busath G, Vazquez A. *Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure*. Psychother. Res. 2012; 22: 638–647.

6. Crits-Christoph P, Ring-Kurtz S, Hamilton JL, Lambert MJ, Gallop R, McClure B. et al. *A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs*. J. Subst. Abuse Treat. 2012; 42: 301–309.

7. Simon W, Lambert MJ, Busath G, Vazquez A, Berkeljon A, Hyer K. et al. *Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomized controlled study*. Psychother. Res. 2013; 23: 287–300.

8. Shimokawa K, Lambert MJ, Smart DW. *Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of psychotherapy quality assurance system*. J. Cons. Clin. Psychol. 2010; 78: 298–311.

9. de Jong K, Nugter MA, Polak MG, Wagenborg JEA, Spinhoven P, Heiser WJ. *The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation*. Clin. Psychol. Psychother. 2007; 14: 288–301.

10. Lambert MJ, Hannöver W, Nisslmüller K, Richard M, Kordy H. *Fragebogen zum ergebnis von psychotherapie: Zur reliabilität und validitat der deutschen ubersetzung des Outcome Questionnaire 45.2*. Z. Klin. Psychol. Psychother. 2002; 31: 40–46.

11. Lo Coco G, Chiappelli M, Bensi L, Gullo S, Prestano C, Lambert MJ. *The factorial structure of the Outcome Questionnair-45: A study with an Italian sample*. Clin. Psychol. Psychother. 2008; 15: 418–423.

12. Wennberg P, Philips B, de Jong K. *The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample*. Psychol. Psychother. Theory Res. Pract. 2010; 83: 325–329.

13. Aleksandrowicz JW, Bierzyński K, Filipiak J, Kowalczyk E, Martyniak J, Mahoń S. et al. *Kwestionariusze objawowe „S" i „O" – narzędzia służące do diagnozy i opisu zaburzeń nerwicowych*. Psychoterapia 1981; 37: 11–27.

14. Sobański JA, Klasa K, Cyranka K, Mielimąka M, Dembińska E, Müldner-Nieckowski Ł. et al. *Effectiveness of intensive psychotherapy in a day hospital evaluated with Neurotic Personality Inventory KON-2006*. Psychiatr. Pol. 2014 [E-pub ahead of print; DOI: 10.12740/psychiatriapolska.pl/online-first/6].

15. Sobański JA, Klasa K, Rutkowski K, Dembińska E, Müldner-Nieckowski Ł. *Profil objawów pacjentów, którzy przerwali psychoterapię kompleksową w dziennym oddziale leczenia nerwic. Badanie retrospektywne*. Psychiatr. Psychoter. 2013: 9(4): 14–30.

16. Sobański JA, Müldner-Nieckowski Ł, Klasa K, Dembińska E, Rutkowski K, Cyranka K. *Traumatic childhood sexual events and secondary sexual heath complaints in neurotic disorders*. Arch. Psychiatry Psychother. 2013; 15(3): 19–32.

17. Sobański JA, Cyranka K, Rodziński P, Klasa K, Rutkowski K, Dembińska E. et al. *Are neurotic personality traits and neurotic symptoms intensity associated with suicidal thoughts reported by patients of a day hospital for neurotic disorders?* Psychiatr. Pol. 2014 [E-pub ahead of print; DOI: 10.12740/psychiatriapolska.pl/online-first/5].

18. Bludworth JL, Tracey TJG, Glidden-Tracey C. *The bi-level structure of the Outcome Questionnaire-45*. Psychol. Assess. 2010; 22: 350–355.

19. Kim SH, Beretvas SN, Sherry AR. *A validation of the factor structure of OQ-45 scores using factor mixture modeling*. Meas. Eval. Couns. Dev. 2010; 42: 275–295.

20. Minami T, Davies DR, Tierney SC, Bettmann JE, McAward SM, Averill LA. et al. *Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center*. J. Couns. Psychol. 2009; 56: 309–320.

21. Mueller RM, Lambert MJ, Burlingame GM. *Construct validity of the outcome questionnaire: A confirmatory factor analysis*. J. Pers. Assess. 1998; 70: 248–262.

22. Lambert MJ, Morton JJ, Hatfield D, Harmon C, Hamilton S, Reid RC. et al. *Administration and Scoring Manual for the Outcome Questionnaire – 45*. Orem, UT: American Professional Credentialing Services; 2004.

23. Derogatis LR. *The SCL-90 manual: Scoring, administration and procedures for the SCL-90*. Baltimore: Johns-Hopkins University School of Medicine, Clinical Psychometrics Unit; 1977.

24. Beck AT, Steer RA, Garbin MG. *Psychometric properties of the Beck Depression Inventory: Twenty-five years later*. Clin. Psychol. Rev. 1988; 8: 77–100.

25. Spielberger CD. *Manual for the State-Trait Anxiety Inventory: STAI (FormY)*. Palo Alto, CA: Consulting Psychologists Press; 1983.

26. Timmerman M, Lorenzo-Seva U. *Dimensionality assessment of ordered polytomous items with parallel analysis*. Psychol. Methods 2011; 16: 209–220.

27. Hooper D, Coughlan J, Mullen M. *Structural equation modelling: Guidelines for determining model fit*. Electron. J. Business Res. Methods 2008; 6: 53–60.

28. Osborne JW, Costello AB, Kellow JT. *Best practices in exploratory factor analysis*. In: Osborne JW. ed. *Best practices in quantitative methods*. Newbury Park, CA: Sage; 2008. p. 86–124.

29. Jacobson NS, Truax P. *Clinical significance: A statistical approach to defining meaningful change in psychotherapy research*. J. Cons. Clin. Psychol. 1991; 59: 12–19.

30. Lambert MJ, Hansen NB, Bauer S. *Assessing the clinical significance of outcome results*. In: Nezu AM, Maguth-Nezu C. ed. *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychological interventions*. New York: Oxford University Press; 2008. p. 359–378.

31. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.

32. Zalewska M, Niemiro W, Samoliński B. *MCMC imputation in autologistic model*. Monte Carlo Methods and Appl. 2010; 16: 421–438.

33. Zalewski T. *Inne Stany. Czym różnią się od nas Amerykanie?* Warsaw: Biblioteka „Polityki"; 2011.

34. Kahn WA. *Psychological conditions of personal engagement and disengagement at work*. Acad. Manage. J. 1990; 4: 692–724.

35. Ayduk O, Mendoza-Denton R, Mischel W, Downey G, Peake PK, Rodriguez M. *Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity*. J. Pers. Soc. Psychol. 2000; 5: 776–792.

36. Mannheim K. *The conditions of social performance*. London: Routledge; 2002.

37. Marsh HW, Hau K, Grayson D. *Goodness of fit in structural equation models*. In: Maydeu – Olivares A, McArdle JJ. ed. *Contemporary psychometrics: A festschrift for Roderick P. McDonald*. Mahwah, NJ: Erlbaum; 2005. p. 275–340.

38. Śliwka P, Simon W. *Markov Chains as a Tool Measuring Effectiveness of a Psychotherapy Process*. In: BIOMAT 2014. Proceedings of the International Symposium on Mathematical and Computational Biology, edited by: Rubem P Mondaini, p. 232-244, DOI: 10.1142/9789814667944_0015, 2015

39. Aleksandrowicz JW, Sobański JA. *Symptom checklist S-III*. Psychiatr. Pol. 2011; 45(4): 515–526.

Address: Witold Simon
Konsonans, Psychotherapy Centre
02-796 Warszawa, Wąwozowa Street 22/61