

VALIDATION OF THE OVERALL DEPRESSION SEVERITY AND IMPAIRMENT SCALE (ODSIS) IN THE CZECH REPUBLIC

PETR MIKOŠKA¹, LUKÁŠ NOVÁK¹, TOMÁŠ BOK¹, MICHAELA LADMANOVÁ¹,
MARTIN FŮLEP¹, RICHARD KOŘINEK¹, LUBOR PILÁRIK²

¹Olomouc University Social Health Institute, Palacký University in Olomouc

²Faculty of Education, Matej Bel University in Banská Bystrica

ABSTRACT

Objectives. The Overall Depression Severity and Impairment Scale (ODSIS) is a brief self-report transdiagnostic measure designed to assess severity and functional impairment associated with depression. Its broader focus makes the measure useful for screening as well as routine outcome monitoring and assessment of overall impact of treatment. The measure is widely used and well applicable in both clinical settings and research context. This study was designed to evaluate the psychometric properties of the Czech version of ODSIS.

Statistical analysis. Ordinal Confirmatory Factor Analysis (CFA) was conducted to explore dimensionality. We also examined the ODSIS items using Item Response Theory (IRT). Moreover, internal consistency, test-retest reliability and construct validity were tested. Finally, the cut-off scores for the Reliable Change Index (RCI) and the Clinically Significant Change Index (CSI) were calculated.

Sample and setting. The present study assessed selected psychometric properties of the Czech version of the ODSIS in three separate samples: a large general representative sample (n=1738), a clinical sample (outpatients and inpatients; n=58) and a separate online retest sample (n=30). In addition to the ODSIS, participants were asked to respond to conventional measures of depression, anxiety, personality traits, self-esteem, life satisfaction and other scales to determine convergent and divergent validity.

Results. Higher depression was observed in females, in widows and widowers, retirees and students. Confirmatory factor analysis indicated an excellent fit of the modified unidimensional model with correlated errors between item 1 and 2: $\chi^2(4) = 8.33$; $p < 0.080$; CFI = 1.000; TLI = 1.000; RMSEA = 0.025, SRMR = 0.008. The ODSIS score was positively associated with guilt and shame, neuroticism, anxiety, perceived stress and an established measuring tool for depression. The ODSIS yielded an excellent internal consistency (Cronbach's $\alpha = 0.94$, McDonald's $\omega = 0.96$), and the temporal stability of the ODSIS score was satisfactory ($r = 0.65$). The ODSIS items had a high discrimination ability and their measurement precision was highest in individuals with a high degree of depression. Differential item functioning revealed that the ODSIS items assess depression in the same way between males and females.

Study limitation. The main limitation is the small clinical sample size, the use of self-report questionnaires for validity testing and lower test-retest reliability.

key words:

ODSIS,
depression,
assessment,
psychometrics,
measurement

INTRODUCTION

Depression is one of the most common mental disorders and is usually associated with feelings of sadness and/or loss of interest in activities that were previously enjoyed (Tiller, 2013). Depression may be associated with significant cognitive, emotional and

Submitted: 16. 12. 2022; L. N., Univerzitní 244/22, 771 11 Olomouc, Czech Republic; e-mail: lukas.novak@oushi.upol.cz

This work was supported by a Palacký University Olomouc Young Researcher Grant, project: The efficacy, effectiveness, and use of Emotion-Focused Therapy in counselling for university students: an experimental study (Contract No. JG 2020 006).

The datasets generated and/or analysed during the current study are available in the Open Science Framework (<https://osf.io/>) repository, <https://osf.io/bn3w8/>

physical problems which can lead to functional impairments (Godard et al., 2012). The overall prevalence of depression in the most recent European study (Torre et al., 2021) was 6.38%, ranging from 2.58% in Czechia to 10.33% in Iceland. In addition, some other studies showed a significant increase in the prevalence of depression during the COVID-19 pandemic (Bueno-Notivol et al., 2021). In 2020, the prevalence of depression tripled compared to November 2017 in Czechia (Winkler et al., 2020), and a later study (Winkler et al., 2021) showed that the Czech population's mental health had not returned to pre-COVID-19 levels. Therefore, depression represents an extraordinary burden on health care systems and a serious problem in recent years (Hidaka, 2012).

In the past, many studies have focused on measuring depressive symptoms and validating scales designed for this purpose. Hence, well-validated instruments, such as the Patient Health Questionnaire: PHQ-9; Kroenke et al. (2002), the Beck Depression Inventory-II: BDI-II; Beck et al. (1996) or the Self-Rating Depression Scale: SDS; Zung (1965), are available. However, the mentioned scales are symptom-focused and may not sufficiently evaluate overall severity and impairment of depression (Bentley et al., 2014; Mira et al., 2019). There is a study in the Czech Republic that examined the psychometric parameters of the Czech abbreviated version of the ODSIS (Sandora et al., 2021). While in our study the original ODSIS (Bentley et al., 2014) is the subject of investigation. The purpose of the Sandora et al. (2021) study was to adapt shortened versions of the ODSIS measure for the purposes of long test batteries where a full version of the ODSIS would increase response burden of participants. However, in our study we adapted a full version of the ODSIS. Therefore, no data collected by Sandora were used in the present study.

In response to the limitations of the above-mentioned screening instruments, the Overall Depression Severity and Impairment Scale was developed (ODSIS; Bentley et al. (2014). It is a five-item measure for assessing the severity and functional impairment of depression, more precisely its frequency and intensity, and also its interference with school/work and interpersonal relationships in the past week (Bentley et al., 2014). Importantly, previous validation studies of the ODSIS supported its very good psychometric parameters (Bentley et al., 2014; Ito et al., 2015; Mira et al., 2019; Osma et al., 2019). For instance, Ito et al. (2015) present value of Cronbach's alpha coefficient of 0.96, Osma et al. (2019) then a value of 0.94. The ODSIS also shows promising values (ICC = 0.75 for non-clinical sample; ICC = 0.73 for clinical sample with following diagnoses: major depressive disorder, panic disorder, social anxiety disorder and obsessive-compulsive disorder) of retest reliability within a two months interval (Ito et al., 2015). In some previous studies Mira et al. (2019) was reported very good convergent validity of the ODSIS with established tools (i.e. PHQ-9 or BDI-II) for measuring depression. The same studies showed adequate discriminant validity of the ODSIS with pre-existing measures of anxiety, quality of life or emotion regulation.

The psychometric properties of the ODSIS have mainly been explored in English speaking countries and the Eastern Asia (Bentley et al., 2014; Mira et al., 2019; Osma et al., 2019). Therefore, we aimed to examine its psychometric parameters in Czech language and to provide further evidence for transcultural validity of the scale. Specifically, the goal of the present study was to examine its psychometric qualities in the Czech Republic, because a short tool for assessing the severity and functional impairment due to depression has not yet been validated in this country. The specific objectives were (1) examination of the selected psychometric parameters (i.e., the unidimensional factor structure, internal consistency, test-retest reliability, convergent and divergent validity) of the ODSIS in a Czech sample and testing differential item functioning; (2) establishing a cut-off score for the reliable change index (RCI) and

the clinically significant change index (CSI); and (3) establishing the Czech percentile norms for the ODSIS. The datasets generated and/or analysed during this study, including percentile norms, are available in the Open Science Framework repository (<https://osf.io/bn3w8/>).

METHODS

Participants

The study was approved by the Ethics Committee of the Olomouc University Social Health Institute, Palacky University Olomouc (No. 2021/6). Participants were required to provide informed consent prior to administration of the surveys. They could quit the surveys at any time without providing a reason. Subjects were informed about what the goals of the present study are and also about the fully voluntary nature of their participation. They were also told that the survey is anonymous.

To explore the psychometric properties of the ODSIS, data from three separate samples were collected. A quality check and outlier screening were performed for each sample. In sample 2 and 3, we additionally removed respondents, who responded incongruently to control questions regarding their height and weight. The outlier screening was done using Median Absolute Deviation (MAD). The MAD function was used to identify participants who answered a very large number of survey questions in the same way. Such respondents were consequently removed.

Sample 1

Participants were Czech adults from a nationally representative sample that was created based on a stratified random sampling method. Specifically, participants were stratified based on age, region of origin and sex. Representativeness was derived from the 2019 baseline sample of the adult population of the Czech Republic (<https://www.czso.cz/csu/czso/vekove-slozeni-obyvatelstva-2019>). In terms of relative frequencies, there was no deviation of the sample from the base sample, which means that the survey is representative of the adult population of the Czech Republic. Data was collected between September 2020 and November 2020. From the initial 2032 participants asked to participate, 1769 (87.1%) agreed with their participation in the study. Of the people who rejected participation, the most frequent reason was a lack of time (38.9%) and an unwillingness to participate (26.3%). Items in the survey were administered to participants by trained specialists ($n = 216$) through the Standard Survey Interviewing method (face-to-face). Subsets of data from this survey were also used by previous studies (Mikoska et al., 2022). Data screening indicated that all participants answered all the survey items. Outlier screening based on MAD suggested that no respondent answered a large number of the survey items in the same way. Participants with age lower than 18 years were excluded ($n = 31$). Therefore, the final number of participants was 1738 (Age: $M = 47.75$, $SD = 17.60$, range: 18–94; Females: 51.38%).

Sample 2

This sample consisted of patients from four psychological outpatient clinics in Nový Jičín and Olomouc and hospitalized patients from the psychiatric clinic of the Hradec Králové University Hospital over the age of 18. Data was collected by psychologists using the paper and pencil method between July and October 2021. Respondents who provided low quality answers to control survey items regarding participants weight and height were excluded ($n = 1$). The tolerance in these two repeatedly asked questions was set as 2 kilograms and 2 centimeters. Outlier screening based on MAD suggested that no respondent answered a large number of the survey items in the

same way. Respondents with age < 18 years ($n = 2$) were also excluded, resulting in 58 subjects (Age: $M = 35.31$, $SD = 14.94$, range: 19 - 82; Females: 70.69%). We used Sample 2 for validity testing. These were patients with various diagnoses (mostly F3 or F4 ICD diagnoses). For the purposes of testing the cut off score, however, we only worked with depression diagnoses. A list of presented mental health issues can be found in the Table 1.

Table 1 Mental diagnoses of patients in the Sample 2 ($n = 58$)

Diagnosis	n (%)
Affective (mood) disorders (F30-39)	15 (23.81)
Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders (F40-48)	33 (52.38)
Disorders of adult personality and behaviour (F60-69)	4 (6.35)
Others	4 (6.35)
Not filled in	7 (11.11)

Note. n = number of respondents; diagnoses are based on ICD-10 diagnostic system

Sample 3

To explore test retest reliability, an online sample of Czech adults was collected between September and November 2021. Data was obtained using a convenience sampling method. From the initial 61 participants, we excluded those who did not complete the survey a second time ($n = 30$), resulting in 31 subjects. Outlier screening based on MAD suggested that no respondent answered a large number of the survey items in the same way. Furthermore, participants responding incongruently to control questions (which were the same as in the Sample 2) were also excluded ($n = 1$), resulting in 30 participants (Age: $M = 40.05$, $SD = 10.82$, range: 20 - 61; Females: 63.33%). A more detailed description of the three samples can be found in in Table 2.

Measures

The Overall Depression Severity and Impairment Scale: ODSIS; Bentley et al. (2014) is a 5-item questionnaire which captures the severity and impairment of depressive symptomatology. The questionnaire assesses the most important domains of depression and subsyndromal depressive symptoms while remaining independent of a particular mood disorder subtype, as delineated by DSM-5 (e.g., persistent depressive disorder, premenstrual dysphoric disorder; APA (2013)). The items ask about the frequency and intensity of depression, the degree in which depression impacts engagement and interest in activities and the extent of interference of depression with necessary daily functioning, social life and relationships. All items are rated on a 5-point Likert scale ranging from 0 to 4, where 0 = “None” and 4 = “Extreme”. Higher scores indicate greater depression severity and impairment. For this study, the Czech translation of the original ODSIS questionnaire was used. The Czech version of the ODSIS is available in the corresponding author on request and only for personal or research purposes.

The translation process followed the WHO (2016) guidelines i.e., forward and backward translation, expert panel, qualitative interview. To assure a proper understanding of the term “depression” for all participants, a definition of this construct was administered together with the questionnaire items.

The Overall Anxiety Severity and Impairment Scale: OASIS; Norman et al. (2006) is a short five-item self-rating scale that assesses the severity of anxiety disorders. Its advantages include that it is not burdensome for respondents, it is easy to

Table 2 Socio-demographic characteristics of the study samples

Characteristic	Sample 1 N = 1,738	Sample 2 N = 58	Sample 3 N = 30
Gender			
Male	845 (48.62%)	17 (29.31%)	10 (34.48%)
Female	893 (51.38%)	41 (70.69%)	19 (65.52%)
Family status			
In partnership/married	875 (50.35%)	28 (49.12%)	23 (79.31%)
Not married	460 (26.47%)	22 (38.60%)	4 (13.79%)
Divorced	222 (12.77%)	7 (12.28%)	2 (6.90%)
Widow/widower	181 (10.41%)		
Education			
Basic school	118 (6.79%)	2 (3.45%)	
Vocational school or without graduation high school	512 (29.46%)	9 (15.52%)	
High school or higher vocational school	714 (41.08%)	24 (41.38%)	
University	394 (22.67%)	23 (39.66%)	
Economical status			
Intellectual worker (medical doctor)	311 (17.89%)		
Entrepreneur > 3 employees	35 (2.01%)		
Manual worker	234 (13.46%)		
Unemployed	73 (4.20%)	4 (7.55%)	5 (17.24%)
Employee in services	359 (20.66%)		
Retired	366 (21.06%)	4 (7.55%)	1 (3.45%)
Administrative worker	161 (9.26%)		
Entrepreneur < 3 employees	73 (4.20%)		
Student	126 (7.25%)	11 (20.75%)	1 (3.45%)
Pensioner		6 (11.32%)	
Employee unspecified		24 (45.28%)	14 (48.28%)
Entrepreneur		4 (7.55%)	8 (27.59%)
Religiosity			
Religious, member of a church	187 (10.76%)	6 (10.53%)	
Religious, not member of a church	414 (23.82%)	19 (33.33%)	
Not religious	1,137 (65.42%)	32 (56.14%)	
Income of family (in Czech Crowns)			
< 10 000	20 (1.15%)		
10.001-20.000	259 (14.90%)		
20.001-30.000	380 (21.86%)		
30.001-40.000	372 (21.40%)		
40.001-50.000	281 (16.17%)		
50.001-60.000	213 (12.26%)		
60.001-70.000	110 (6.33%)		
70.001+	103 (5.93%)		

score, and it also takes into account the respondent's functional impairments associated with the level of anxiety experienced. Its factorial validity and superior psychometric properties have been demonstrated several times. In this study the internal consistency of the OASIS was excellent with Cronbach's alpha: 0.9, 95% CI[0.86-0.94] and McDonald's omega: 0.95.

The Patient Health Questionnaire: PHQ-9; Kroenke et al. (2001); is a brief measure assessing the severity of depression. Its nine items reflect the nine DSM-IV (APA, n.d.) depression diagnostic criteria, thus, asking about the frequency of various symptoms (e.g., dropped mood, sleeping difficulties, changes in appetite) and the presence and duration of suicide ideation. The items are rated on a 4-point Likert scale, where 0 = "Not at all" and 3 = "Nearly every day". Higher scores indicate higher severity of depression. The questionnaire also incorporates a non-scored question examining the degree to which depressive symptoms have affected a patient's level of functioning. The Czech translation of the PHQ-9 (Daňsová et al., 2016) has two factors – labeled as "somatic symptoms" and "affective-cognitive symptoms" – and an acceptable internal consistency (Cronbach's $\alpha = 0.79$). The measure positively correlates with the widely used Beck Depression Inventory-II (Beck et al., 1996). Therefore, the Czech version can be considered as a reliable and valid tool for assessing depression in this population. In our study the internal consistency of the PHQ-9 was good with Cronbach's alpha: 0.9, 95% CI[0.85-0.93] and McDonald's omega: 0.93.

The Generalized Anxiety Disorder Questionnaire: GAD-7; Spitzer et al. (2006); is a 7-item measure assessing the symptoms of generalised anxiety disorder, as defined by DSM-IV (APA, n.d.). The items examine the frequency of feeling nervous or anxious, the inability to control thoughts, difficulties with relaxing, perceived restlessness, etc. They are answered using a 4-point Likert scale, where 0 = "Not at all" and 3 = "Nearly every day". Higher scores indicate greater severity of generalised anxiety disorder. The questionnaire also includes a non-scored item assessing the degree to which anxiety symptoms have affected a patient's level of functioning. The Czech version of the GAD-7 has good psychometric properties (Prikner, 2021). Its internal consistency was good across samples (Cronbach's $\alpha = 0.73-0.92$) in the mentioned study. Test-retest reliability between four administrations reaches mean values (ICC = 0.54). In our study the internal consistency of the GAD-7 was good with Cronbach's alpha: 0.89, 95% CI[0.84-0.93] and McDonald's omega: 0.94.

The Rosenberg's Self-Esteem Scale: RSES; Rosenberg (1965); is a 10-item measure evaluating individual self-worth by measuring both positive and negative feelings about the self. The items are rated on a 4-point Likert scale ranging from "Strongly agree" to "Strongly disagree". Higher scores indicate higher self-esteem. The Czech translation of the scale (Blatný & Osecká, 1994) is believed to have a one-dimensional structure and good internal consistency. In our study the internal consistency of the RSES was excellent with Cronbach's alpha: 0.9, 95% CI[0.85-0.93] and McDonald's omega: 0.92.

The Big Five Inventory: BFI; John and Srivastava (1999); is a 44-item measure which captures five personality traits based on the Five-factor model of personality. These are Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness. Each of the factors is then divided into six personality facets. For the purposes of this study, only three factors were used – Agreeableness, Neuroticism, and Openness. Their items consist of statements about the patient which are answered using a 5-point Likert scale ranging from 1 to 5, where 1 = "Disagree strongly" and 5 = "Agree strongly". Higher scores reflect higher agreeableness, neuroticism and openness. The Czech version (Hřebíčková & Urbánek, 2001) supports the factor structure and shows

good internal consistency of the three above-mentioned subscales (Cronbach's $\alpha = 0.65-0.83$). In our study Cronbach's alpha and McDonald's omega for the N in the clinical sample are 0.83, 95% CI[0.75-0.89] and 0.89, for O: 0.87, 95% CI[0.81-0.91] and 0.9 and for A: 0.44, 95% CI[0.18-0.63] and 0.73.

The Perceived Stress Scale-10: PSS-10; Cohen et al. (1983); is a 10-items stress assessment questionnaire. The items ask about the frequency of perceived stress in various situations or about the coping skills related to these situations. The questions are answered on a 5-point Likert scale, where 0 = "Never" and 4 = "Very often". Higher score indicates a higher level of perceived stress. The Czech translation (Buršíková Brabcová & Kohout, 2018) has good internal consistency (Cronbach's $\alpha = 0.87$). In our study the internal consistency of the PSS-10 was good with Cronbach's alpha: 0.87, 95% CI[0.82-0.92] and McDonald's omega: 0.91.

The Guilt and Shame Experience Scale: GSES; Maliňáková et al. (2019); is an 8-item questionnaire capturing the experiences of guilt and shame. The items are answered on a 4-point Likert scale, where 1 = "Not at all" and 4 = "Significantly". Higher scores imply more experiences of guilt and/or shame. The psychometric evaluation revealed excellent internal consistency (Cronbach's $\alpha = 0.85$ for the 8-item version). In our study the internal consistency of the GSES was good with Cronbach's alpha: 0.86, 95% CI[0.8-0.91] and McDonald's omega: 0.91.

The Self-Compassion Scale: SCS; Neff (2003); is a 26-item questionnaire assessing an overall level of self-compassion. The items are rated on a 5-point Likert scale ranging from 1 to 5, where 1 = "Almost never" and 5 = "Almost always". Higher scores reflect greater self-compassion. The Czech translation (Benda & Reichova, 2016) has an excellent internal consistency (Cronbach's alpha $\alpha = 0.89$). In our study the internal consistency of the SCS was excellent with Cronbach's alpha: 0.91, 95% CI[0.87-0.94] and McDonald's omega: 0.94.

The Clinical Outcomes in Routine Evaluation: CORE OM; Evans et al. (2000); is a 34-item pan-theoretical and pan-diagnostic questionnaire of global psychological distress. The measure draws upon the view of practitioners and their perceptions of the most important aspects of psychological well-being. The questionnaire asks about the frequency of problems occurring in four domains – subjective well-being, symptoms, functioning, and risk to the self and others. The items are answered on a 5-point Likert scale, where 0 = "Not at all" and 4 = "Most or all the time". Higher scores indicate greater psychological distress. The Czech translation (Seryjová Juhová et al., 2018) shows good psychometric properties. It has the expected four-factor structure and excellent internal consistency (Cronbach's alpha $\alpha = 0.93$). The scale assessing the risk for the self and others was not included in this study. In our study all of the CORE OM subscales used yielded satisfactory values of Cronbach's alpha and McDonald's omega: well-being: 0.78, 95% CI[0.67-0.86] and 0.82, functioning subscale: 0.85, 95% CI[0.78-0.9] and 0.94 as well as problems subscale: 0.86, 95% CI[0.8-0.91] and 0.93.

The Satisfaction With Life Scale: SWLS; Diener et al. (1985); is a brief questionnaire measuring cognitive judgements about one's life satisfaction. The five items consist of positive statements which are rated on a 7-point Likert scale ranging from 1 = "Strongly disagree" to 7 = "Strongly agree". Higher scores reflect greater satisfaction with life. The Czech version (Lewis et al., 1999) has a one-dimensional structure. The subsequent Czech study (Benda & Reichova, 2016) revealed the good internal consistency (Cronbach's alpha $\alpha = 0.87$) of the SWLS. We also found good internal consistency of the SWLS, with Cronbach's alpha: 0.87, 95% CI[0.81-0.92] and McDonald's omega: 0.91 in our study.

Data analysis

The assumptions of statistical tests were checked in all three samples. Multivariate normality testing indicated that multivariate normality was not achieved in the retest and representative sample, but was achieved in the clinical sample. Thus, in the first two samples, non-parametric methods were used. The Breusch-Pagan test, as well as inspection of the residual plot, suggested heteroscedasticity in the first two samples. Therefore, statistical methods correcting for unequal variances were applied. Exploration of missing data suggested that there is no pattern in the missing data; therefore, missing cases were removed listwise.

Item Response Theory approach

The Item Response Theory (IRT) approach was used to explore the psychometric parameters of the ODSIS (sample 1: $n = 1738$). The monotonicity and local independence were tested using the LD chi-square statistic (Chen & Thissen, 1997) and the method presented by van der Ark (2007). Due to imbalances between response categories, we used the Generalised Partial Credit model. To evaluate model fit within IRT framework C2 index was used. This index is typically used when fitting an IRT model on ordinal data (Cai & Monroe, 2014) and when M2 statistic does not have sufficient number of degrees of freedom. In the first step, a one parameter - constrained model was fitted. This model estimates only the difficulty parameter (b) – i.e. how difficult is an item to endorse (Kim & Wilson, 2020) – while fixing the discrimination parameter (a_i) to be equal. As suggested by Huang et al. (2017), the discrimination parameter indicates how an item discriminates between individuals having the similar degree of a latent trait (θ). In the second step, we estimated an IRT model in which the discrimination parameter was allowed to vary. Consequently, these two models were compared using a likelihood ratio test. The significance of this test indicates the superiority of the more complex model (Blötner & Beisemann, 2022). As indicated by Belous et al. (2021), values of the alpha parameter greater than 2 are considered as best, values above 1 as good and values greater than 0.75 as acceptable. Item information function curves were created from both parameters. Measurement precision of the ODSIS was assessed by a test information function. Differential item functioning was explored using ordinal logistic regression.

Classical Test Theory approach

The unidimensionality of the instrument was tested via ordinal Confirmatory Factor Analysis (CFA) on the representative sample ($n = 1738$). The KMO measure and Bartlett test were used to assess the appropriateness of the correlation matrix for conducting factor analysis. The fit of the model was evaluated using the following indices: the Standardised Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Root Mean Square Error of Approximation (RMSEA). The RMSEA and SRMR values of < 0.08 and < 0.05 represents an acceptable and a good fit, respectively (Hooper et al., 2008; Vandenberg & Lance, 2000). In the CFI and TLI, values > 0.95 and > 0.97 suggest an acceptable and a good fit, respectively (Jackson et al., 2009; Schermelleh Engel et al., 2003). The implied model and observed covariance were compared using the chi-square test. Due to the number of response categories of the instrument (DiStefano & Morgan, 2014), the model was fitted using the Diagonally Weighted Least Squares estimator (DWLS) based on a polychoric correlation matrix. The reliability of the ODSIS in terms of internal consistency was evaluated using McDonald's omega and also by Cronbach's alpha. The values of the Cronbach's alpha and the McDonald's Omega > 0.8 can be interpreted as good (Feißt et al., 2019).

Convergent, divergent and concurrent validity were inspected using the zero-order Spearman correlation coefficient (Sample 2). The test-retest reliability of the ODSIS was explored using the two-way mixed effects intraclass correlation coefficient (Sample 3). The time between the two measurements was seven days. The Jacobson and Truax (1991) formula was used to calculate the cut-off for a clinically significant change (CSC). In this formula SD_0 and M_0 represent standard deviation and mean of the ODSIS of the non-clinical sample. SD_1 and M_1 stand for mean and standard deviation of the ODSIS of the clinical sample. The Jacobson and Truax (1991) equation is as follows:

$$c = \frac{SD_0 M_1 + SD_1 M_0}{SD_0 + SD_1}$$

Moreover, the cut-off value for the Reliable Change Index (RCI), sometimes called the statistical significant change, was also estimated. The formula for calculating RCI presented by Wiger and Solberg (2001) is as follows:

$$RCI = 1.96 \sqrt{2SD^2(1 - rel)}$$

As a representative sample was recruited in this study, population norms were created. These norms were created using polynomial regression, permitting relationships between ODSIS expected score, explanatory variable (age) and norm score to be modelled (Gary et al., 2021). All statistical calculations were conducted in R programming environment (version 4.2.1) primarily using the packages: psychtoolbox (Novak, 2022), usf (Peters, 2021), cNORM (Gary et al., 2021), psych (Revelle, 2021), mirt (Chalmers, 2012), lordif (Choi et al., 2016), lavaan (Rosseel, 2012) and papaja (Aust & Barth, 2020).

RESULTS

Confirmatory Factor Analysis

The KMO and also the Bartlett test ($\chi^2(10) = 7,967.73$, $p < .001$) indicated that data from the first sample are appropriate for factor analysis. The CFA results suggested that the CFI and the TLI yielded excellent values. However, the RMSEA suggested that the fit of the unidimensional model is not optimal: $\chi^2(5) = 111.98$, $p < .001$, CFI=0.999, TLI=0.998, SRMR=0.024, RMSEA=0.111, 90% CI [0.094-0.129]. Inspection of the modification indices suggested that if error terms between items 1 and 2 were allowed to correlate, the model fit might be increased. As the correlation between these error terms is theoretically justifiable (Ito et al., 2015; Osma et al., 2019) and also in line with the approach of the previous validation studies of the ODSIS, the decision was made to let these error terms to correlate.

After the error terms between the items 1 and 2 were allowed to correlate, the unidimensional model had an excellent fit: $\chi^2(4) = 8.33$, $p = 0.080$, CFI=1.000, TLI=0.9999, SRMR=0.008, RMSEA=0.025, 90% CI [0-0.049]. The superior fit of the later model was also supported by the likelihood ratio test: $\chi^2(1) = 84.79$; $p < .001$. In the modification version of the ODSIS model, correlations between errors of manifest variables were very low: $r_{max} = 0.01$. This suggests that the model neither underpredicts nor overpredicts covariance between ODSIS items. This has two following implications: 1) we did not omit another important element that should be included in the model; 2) there does not seem to be a redundant element in this model.

It is important to note that values of the CFI and TLI were very high especially in modified model. It was possible that such high values might be a result of the just-identified model. However, after exploration of the number of degrees of freedom this explanation was rejected. Another explanation might rest on the relatively high association between ODSIS items. This explanation was supported by inspection of the correlation matrix indicating very high correlations between ODSIS items (Table 3).

Item statistic and internal consistency

Factor loadings (λ) in the unidimensional model were very high, ranging from $\lambda = 0.91$ up to $\lambda = 0.96$. Internal consistency of the the ODSIS was excellent in both the representative sample (Cronbach alpha: 0.94, 95% CI[0.94-0.95]; McDonald's omega: 0.96) as well as in clinical sample (Cronbach alpha: 0.9, 95% CI[0.86-0.94]; McDonald's omega: 0.95). The statistics of the ODSIS items are reported in Table 3. Item-total correlations as well as correlations between the ODSIS items were high overall. The lowest value of item-total correlation and factor loading was observed in item 3 and item 1, respectively.

Table 3 Item statistic and polychoric correlation matrix of the ODSIS items (n = 1738)

Variables	ODSIS_1	ODSIS_2	ODSIS_3	ODSIS_4	ODSIS_5	M	SD	ITC	Skewness	Kurtosis	Factor loadings
ODSIS_1	1					0.38	0.7	0.88	2.15	5.16	0.91
ODSIS_2	0.94** *	1				0.34	0.67	0.92	2.33	5.92	0.94
ODSIS_3	0.86** *	0.88** *	1			0.34	0.65	0.84	2.22	5.50	0.94
ODSIS_4	0.86** *	0.9** *	0.89** *	1		0.28	0.62	0.86	2.65	7.84	0.96
ODSIS_5	0.87** *	0.88** *	0.87** *	0.89** *	1	0.27	0.61	0.85	2.66	7.75	0.94

Note. *** $p < 0.001$, M = Mean, SD = Standard Deviation, ITC = Item-total correlation with correction for item overlap and the scale reliability

Construct validity testing, cut-off scores and test-retest reliability

Spearman correlation coefficient revealed that the ODSIS score was strongly and positively correlated with established depression measure, perceived stress, and anxiety: Table 4 online (<https://osf.io/82qyh>). In addition, ODSIS was strongly and negatively related with self-esteem, well-being, overall functioning and self-compassion. Test-retest reliability examination indicated that temporal stability of the ODSIS was satisfactory: $r = 0.65$, 95% CI [0.30 - 0.85], $p < .001$.

Reliable Change Index, Clinically significant change and cut-off scores

Results revealed that the ODSIS cut-off score (CSC) discriminating the non-clinical from the clinical population is 4.52. Evaluation of the cut-off score for statistically significant change (RCI) indicated that the optimal value is 4.79. To evaluate the cut-off value of the CSC as determined by the equation defined in the Data analysis section, we used sensitivity and specificity analysis. We additionally calculated Area Under the Curve (AUC) values.

As indicated by Receiver Operating Characteristic (ROC – Figure 1), values of area under the curve (AUC) reached 0.92 95% CI [0.87 – 0.97] when discriminating between representative sample (Sample 1) and subset of clinical sample having depression (Sample 2) using the ODSIS. If this rounded CSC cut-off (5) would be used, then sensitivity would reach 0.73 and specificity 0.16 values. This means that, if the above mentioned cut-off is used, then 73% of participants having depression would be correctly classified.

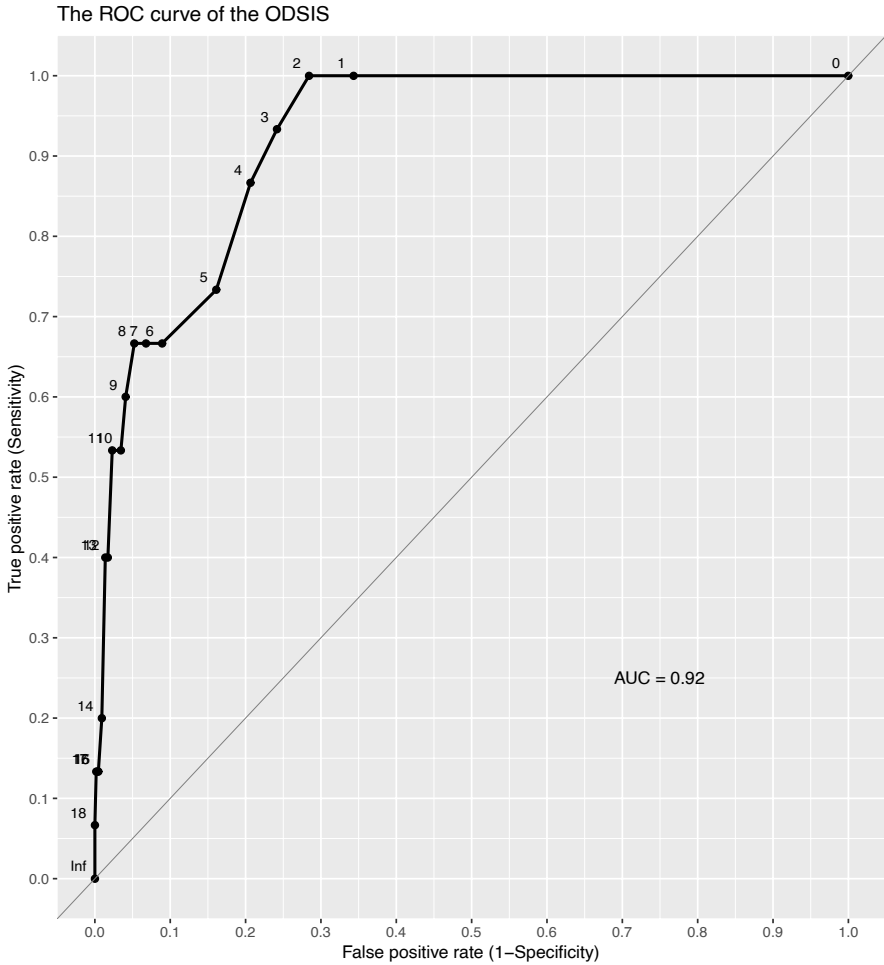


Figure 1 represents the ROC curve depicting the ability of the ODSIS score to differentiate between the clinical sample (Sample 2) and representative sample (Sample 1)

Note. ODSIS = Overall Depression Severity and Impairment Scale, ROC = Receiver Operating Characteristic Curve, black dots inside the figure denote the ODSIS score.

IRT analysis of the ODSIS items (Sample 1)

The testing of locale independence suggested a higher association between the residuals of items 1 and 2. However, the degree of this association was still in the acceptable range. It was further indicated that the monotonicity assumption holds. The testing of item difficulty using a one-parameter model (1PL) revealed that values of

the b_i were located mostly above average of latent depression. This suggests that the ODSIS items were rather difficult for participants. When comparing difficulty across items, it was observed that the ODSIS items are relatively similar in their difficulty (Figure 4 in appendix). However, despite this similarity, item 5 ($b_i = 2.599$) was found to be the most difficult item. The fit of this one-parameter model was acceptable: $M2 = 146.15$, $df = 9$, $p < .001$, $RMSEA = 0.094$ 95% CI [0.081, 0.107], $SRMR = 0.057$, $TLI = 0.985$, $CFI = 0.986$.

In the next step, we created a two-parameter logistic model (2PL) in which the α_i parameter was allowed to vary across the ODSIS items. Consequently, this model was compared with the 1PL model. The likelihood ratio test revealed that the 2PL model is superior to the 1PL model in terms of model fit: $\chi^2(4) = 50.13$, $p < .001$. The slopes of the ODSIS items were very high, suggesting that these items discriminate participants well in degree of depression. The highest discrimination parameter was observed in item 2 ($\alpha_i = 8.346$). Table 5 in appendix presents the frequency of responses to the individual ODSIS items. This table also presents the exact values of difficulty and discrimination parameters. The fit of this two-parameter model was suboptimal: $M2 = 114.44$, $df = 5$, $p < .001$, $RMSEA = 0.112$ 95% CI [0.095, 0.131], $SRMR = 0.051$, $TLI = 0.978$, $CFI = 0.989$.

Item discrimination and the difficulty parameter were used to estimate measurement precision of the ODSIS items. The results are presented in Figure 2. They reveal that item 2 was measuring with the highest precision of all the ODSIS items and that all ODSIS items were measuring with the different precision in approximately the same location of depression (θ).

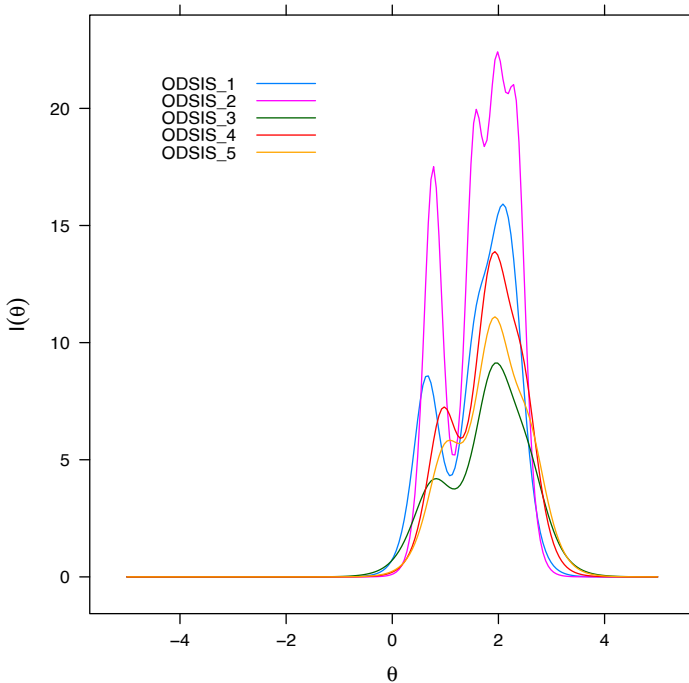


Figure 2 Measurement precision of individual ODSIS items

Note. The y-axis reflects the degree of information/measurement precision ($I(\theta)$). The x-axis represents the level of latent depression (θ).

IRT on the scale level and Differential Item Functioning

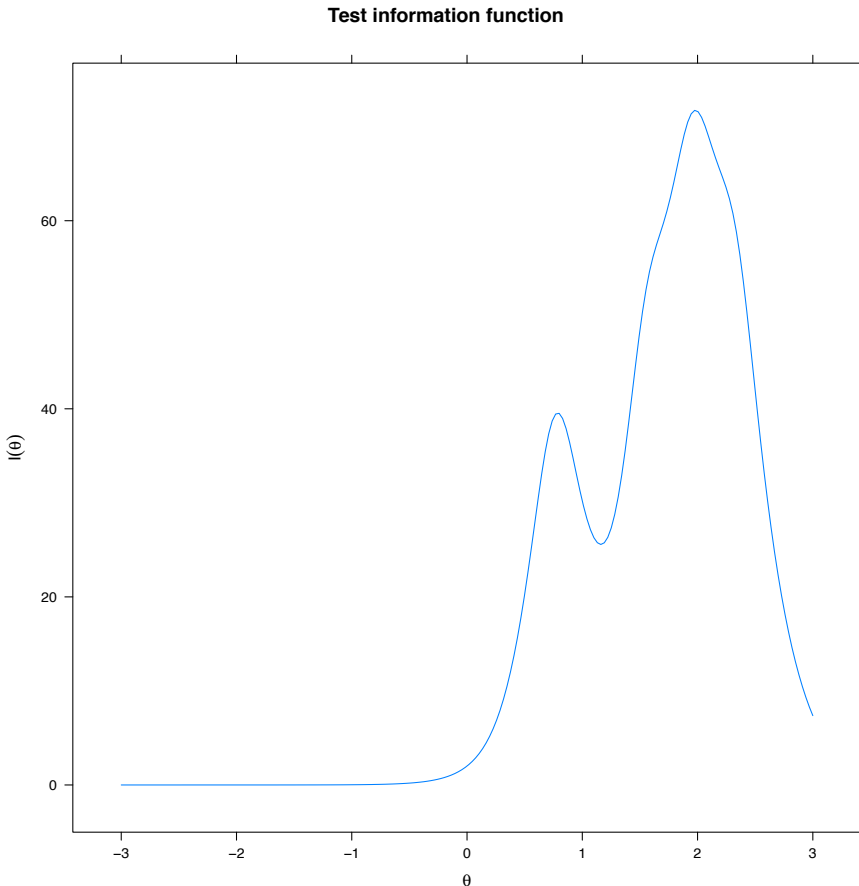


Figure 3 Test information function curve

Note. The x-axis reflects the level of depression (as a latent variable - θ) standardized in Z-scores. The y-axis refers to degree of precision/information obtained - $I(\theta)$ - from measuring different parts of the latent depression.

Figure 3 represents the test information function of the ODSIS. The ODSIS was found to assess depression with very low precision in individuals with low levels of depression. The highest measurement precision is reached in individuals with an above average degree of depression, i.e. with a score higher than two standard deviations from the mean. However, due to use of the study sample primarily consisting of individuals not having depression (Sample 1), these results should be interpreted with caution. Testing of the differential item functioning revealed that the ODSIS does not include items which would measure differently in males and females. This finding indicates that scores between females and males can be meaningfully compared. Moreover, results of the DIF testing suggest that if future studies will detect a significant difference between males and females in ODSIS score, these differences are more likely reflecting a different levels of state depression rather than differences in how males and females understand questionnaire items.

DISCUSSION

This study aimed to explore the psychometric parameters of the Overall Depression Severity and Impairment Scale (ODSIS) in the Czech Republic. The factor analysis supported a modified-unidimensional solution of the ODSIS. The convergent and discriminant validity was supported by findings that a higher ODSIS score was positively associated with an established depression measure, perceived stress and anxiety and was negatively associated with self-esteem, well-being, overall functioning and self-compassion. The test-retest reliability of the ODSIS was acceptable. IRT analysis revealed that the items discriminated the participants well in their degree of depression severity and impairment, and also revealed that ODSIS items measure with high precision in participants with an above average degree of depression. The assessment of differential item functioning indicated that the items measure depression severity and impairment in the same way in both males and females.

Structure and items

The results of the CFA supported a modified-unidimensional model of the ODSIS. These results are congruent with findings of previous studies (Ito et al., 2015; Osma et al., 2019; Sandora et al., 2021). Moreover, as in previous studies, correlation of residuals between item 1 and 2 was allowed. As this modification was needed across different samples from different cultural environments, it is very likely that the cause of this correlation rests on the instrument construction rather than on factors associated with the study sample. The reason for correlation of error terms between items 1 and 2 might be as follows. In contrast with the remaining items, which assess the degree of impairment resulting from depression, items 1 and 2 assess the frequency and intensity of depression. Therefore, it is possible that the structure of the ODSIS is, in fact, two-dimensional: the first dimension assesses depressive feelings and the second assesses the impairment resulting from depression. This might explain residual correlations between the first two ODSIS items. It is important to note that in the representative sample, correlations between the ODSIS items were very high. This is likely to be caused by large homogeneity of participant responses: most participants selected either the first answer (0) or the second answer (1) on the provided scale. For this reason, test internal consistency of the ODSIS could be overestimated to some degree.

The ODSIS items were found to provide the most information about individuals suffering from a higher degree of depression severity and impairment. To the best of our knowledge only a single study has previously examined the psychometric properties of the ODSIS using the IRT (Bentley et al., 2014). Our results are in line with this study in terms of item difficulty, which was relatively high in both studies. However, in our study overall item difficulty seemed to be slightly higher in comparison to the study of Bentley et al. (2014). For instance, in item 4, an individual having an above average degree of depression (i.e., having a latent depression score approximately one standard deviation above the mean) has a 50% chance of selecting answer two (i.e., the depression caused some interference at work/home/school) over answer one (i.e., no interference at work/home/school from depression). A possible explanation for the higher difficulty of the ODSIS items in our study might lie in the fact that the term “depression” may be understood differently across different language environments. The ODSIS instruction uses the term “depression”, which can cause some confusions for participants about the semantic meaning of that term. There are various denotations such as mood, a symptom or psychiatric disorder. There could be a broad spectrum of mixed connotations such as emotional experience and social consequences of ex-

perienced mood (Marsella, 2003). Although the possibility of multiple interpretations is minimized by the introductory instruction specifying what is meant by the term *depression*, there is a possibility that participants could assess their emotional mood in a more holistic manner of impairment, which explains the large intercorrelation between items with only weak differentiation of depression severity and its functional impact. Another possible explanation for the higher difficulty of the ODSIS items is that the impact of depression on personal life and relationships may occur with a time delay after the individual experiences symptoms of depression.

Convergent and Discriminant Validity

The convergent validity of the ODSIS was supported by a large positive correlation with a well-established depression measure, the Patient Health Questionnaire-9 (PHQ-9). Previous studies on American students (Bentley et al., 2014), Japanese adult patients and out-patients (Ito et al., 2015) and Spanish clinical out-patients (Mira et al., 2019; Osma et al., 2019) found similar large positive correlations with depression measures.

On the other hand, we found medium to large positive correlations of the ODSIS with neuroticism (BFI-N), anxiety (GAD-7, OASIS), stress (PSS-10). These correlations were in lower magnitude than the correlation with depression (PHQ-9). Our findings are in line with the suggestion of Ito et al. (2015) that the ODSIS distinguishes between scores of self-reported depression and anxiety, other somatic symptoms and stress level. The ODSIS score was also independent from agreeableness and openness. This findings support discriminant validity of the ODSIS as was found in previous studies (Bentley et al., 2014; Mira et al., 2019).

The ODSIS was in large negative correlation with outcome measures, such as the Satisfaction with Life Scale (SWLS) or the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) subscales. Previous studies found similar negative correlations of the ODSIS with the SWLS (Ito et al., 2015), quality of life (Mira et al., 2019; Osma et al., 2019) maladjustment (Osma et al., 2019) and mental health (Bentley et al., 2014). This result supports the ability of the ODSIS to detect functional impairment in daily life. In the present study, we found a large negative correlation between self-compassion (SCS) and the ODSIS. To the best of our knowledge, the present study is the first to examine the relationship between the ODSIS and the SCS. Our results are in line with the theoretical assumptions of Neff (2003), who proposed that self-compassion should be positively associated with mental health outcomes, such as fewer symptoms of depression or anxiety and greater life satisfaction. Similarly, Hoge et al. (2013) proposed that an accepting attitude toward one's affect can attenuate distress. The ODSIS was negatively correlated with self-esteem (RSES) and positively correlated with experiencing guilt and shame (GSES). Low self-esteem and shame pertains to a painful focus on the self and guilt as an interpersonally driven emotion arising from the belief that one has hurt another and are associated with depression (Gambin & Sharp, 2018). Overall, the results obtained in this study support the construct validity of the ODSIS.

Reliability

The internal consistency of the ODSIS items was excellent in the Czech sample, consistent with previous validation studies (Ito et al., 2015; Mira et al., 2019). These findings across countries support the unidimensional structure of ODSIS. On the other hand, test-retest reliability in the Czech sample was acceptable, but lower in comparison with previous studies. Only two previous studies tested test-retest reliability (Ito

et al., 2015; Sandora et al., 2021) and found higher test-retest reliability (0.73–0.85). Both studies used an abbreviated version of ODSIS that could to some extent explain differences between studies results. Depression severity and impairment is prone to change depending on the external circumstances, too. Data were collected during the COVID-19 pandemic that could cause greater disturbances in experiences of depression severity and impairment of depression. We cannot conclude that the lower test-retest reliability is attributable to the participants' depression severity and impairment instability or the ODSIS reliability.

Cut-off scores

The findings of this study suggest that the threshold of a 4.52 cut-off sum score (CSC) on the ODSIS discriminates the non-clinical from the clinical population. To prevent pathologizing of the respondents, the value was rounded up (i.e., 5). ROC analysis revealed that the ODSIS value of 5 discriminates between the representative sample and the clinical sample having depression. The cut-off sum score of 5 could be used to screen individuals with depression symptoms for intervention in the Czech Republic. A similar cut-off score (5) was found in the previous study of Mira et al. (2019) on a Spanish population. In contrast, earlier studies found higher cut-offs, such as cut-offs of 8 (Bentley et al., 2014), 11 (Ito et al., 2015) and 10 (Osma et al., 2019). This inconsistency could be due to cultural differences in depression prevalence (e.g., US, Japan, etc.), the COVID-19 pandemic occurring during data collection, or various criteria used in previous studies for clinical sample selection (e.g. major depressive disorder diagnosis; score on BDI-II., etc.). Evaluation of the cut-off score for evaluation of the cut-off score for reliable change index (RCI) indicated that the optimal value is 4.79. The threshold of 5 could thus be used as a criterion of statistical significant change in depression severity and impairment (e.g., through an intervention process). Previous studies did not explore this index, so future research is needed to replicate these findings.

Implications for practice

The Czech translation of the ODSIS represents a brief reliable and valid instrument for assessing depression related severity and impairment. Compared to already existing depression scales focused mainly on the severity or frequency of individual symptoms, the ODSIS additionally captures their impact on a patient's daily functioning, interest and engagement in activities, and interpersonal relationships. This broader focus makes the measure more useful not only for screening but also routine monitoring of therapeutic changes or overall impact of treatment. The results further revealed that the ODSIS assesses depression severity and impairment with relatively high precision in individuals having an above average degree of depression severity and impairment. Specifically, the highest precision was achieved in individuals having the depression score higher than two standard deviations from the mean.

Strengths and Limitations

The first strength of this study is that a representative sample of Czech population was used. Second, the convergent validity of the ODSIS was confirmed with the use of a broad range of measures of psychopathology and well-being. Third, due to its excellent psychometric properties and ease of completion, the Czech version of the ODSIS is suitable for screening use in primary and secondary care or for continuous mapping of a client's depression severity and impairment changes during the psychotherapy/counselling process. Furthermore, due to the above-mentioned characteristics and the defined cut-offs for the CSC and the RCI, it is also a useful measure for research pur-

poses (e.g., in outcome studies). This study has limitations that should be mentioned. First, the data collections were done during the COVID-19 pandemic. The increased prevalence of affective disorders and depression disorders specifically during these circumstances could have influenced the identified cut-off scores in the ODSIS and test-retest reliability. Second, the measures used for the ODSIS validity testing were all self-report scales and questionnaires, which are prone to social desirability bias. Third, the internal consistency of the BFI-A was unsatisfactory, which might be the primary reason why we did not find a significant association between depression (ODSIS) and agreeableness (BFI-A). Fourth, the CSC cut-off scores may have been influenced by the fact that the representative Czech population probably included people who belonged to the clinical population. Other limitations are associated with the need to modify the factor structure of the ODSIS, with the construct validity testing in the smaller clinical sample, and with the assessment of the severity and impairment of depression from the reflective construct point of view.

CONCLUSION

The ODSIS represents a reliable and valid instrument for assessing depression in the adult Czech population. Due to its shortness, excellent psychometric properties, comprehensive percentile norms and cut-off score, it is useful for assessing depression severity and impairment, and therapeutic change in clinical practice.

REFERENCES

- APA (n.d.). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* Washington, DC: American Psychiatric Association.
- APA (2013). *DSM-5 task force diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.
- Ark, L. A. van der (2007). Mokken scale analysis in R. *Journal of Statistical Software*, 20(11), 1-19. <https://doi.org/10.18637/jss.v020.i11>
- Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II: Beck Depression Inventory Manual*. 2nd edn. Psychological Corporation.
- Belous, C. K., Wampler, R. S., & Ledford, B. L. (2021). Testing the gay and lesbian relationship satisfaction scale with item response modeling. *Journal of Couple & Relationship Therapy*, 20(1), 1-14. <https://doi.org/10.1080/15332691.2020.1746460>
- Benda, J., & Reichová, A. (2016). Psychometric characteristics of the Czech version of the Self-Compassion Scale (SCS-CZ). *Československá psychologie*, 60(2), 120-136.
- Bentley, K. H., Gallagher, M. W., Carl, J. R., & Barlow, D. H. (2014). Development and validation of the Overall Depression Severity and Impairment Scale. *Psychological Assessment*, 26(3), 815-830. <https://doi.org/10.1037/a0036216>
- Blatný, M., & Osecká, L. (1994). Rosenbergova škála sebehodnocení: Struktura globálního vztahu k sobě. *Československá psychologie*, 38(6), 481-488.
- Blötner, C., & Beisemann, M. (2022). *The dark triad is dead, long live the dark triad: An item-response theoretical examination of the short dark tetrad*. PsyArXiv. <https://doi.org/10.31234/osf.io/kr352>
- Buršíková Brabcová, D., & Kohout, J. (2018). Psychometrické ověření české verze Škály vnímaného stresu. *E-Psychologie*, 12(1), 37-52. <https://doi.org/10.29364/epsy.311>
- Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data*. National Center for Research on Evaluation, Standards; Student Testing (CREST).
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. <https://doi.org/10.18637/jss.v048.i06>
- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265-289. <https://doi.org/10.2307/1165285>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1-30. <https://doi.org/10.18637/jss.v039.i08>

- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). *Perceived stress scale*. American Psychological Association (APA). <https://doi.org/10.1037/t02889-000>
- Daňšová, P., Masopustová, Z., Hanáčková, V., Kicková, K., & Korábová, I. (2016). Metoda Patient Health Questionnaire-9: Česká verze. [The Patient Health Questionnaire-9: The Czech version]. *Československá psychologie*, 60, 468-481.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13
- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling*, 21(3), 425-438. <https://doi.org/10.1080/10705511.2014.915373>
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health*, 9, 247-255. <https://doi.org/10.1080/713680250>
- Feißt, M., Hennigs, A., Heil, J., Moosbrugger, H., Kelava, A., Stolpner, I., Kieser, M., & Rauch, G. (2019). Refining scores based on patient reported outcomes – statistical and medical perspectives. *BMC Medical Research Methodology*, 19(1), 167. <https://doi.org/10.1186/s12874-019-0806-9>
- Gambin, M., & Sharp, C. (2018). Relations between empathy and anxiety dimensions in inpatient adolescents. *Anxiety, Stress, & Coping*, 31(4), 447-458. <https://doi.org/10.1080/10615806.2018.1475868>
- Gary, S., Lenhard, W., & Lenhard, A. (2021). Modelling norm scores with the cNORM package in R. *Psych*, 3(33), 501-521. <https://doi.org/10.3390/psych3030033>
- Godard, J., Baruch, P., Grondin, S., & Lafleur, M. F. (2012). Psychosocial and neurocognitive functioning in unipolar and bipolar depression: A 12-month prospective study. *Psychiatry Research*, 196(1), 145-153. <https://doi.org/10.1016/j.psychres.2011.09.013>
- Hidaka, B. H. (2012). Depression as a disease of modernity: Explanations for increasing prevalence. *Journal of Affective Disorders*, 140(3), 205-214. <https://doi.org/10.1016/j.jad.2011.12.036>
- Hoge, E. A., Bui, E., Marques, L., Metcalf, C. A., Morris, L. K., Robinaugh, D. J., Worthington, J., Pollak, M., & Simon, N. M. (2013). Randomized controlled trial of mindfulness meditation for generalized anxiety disorder: Effects on anxiety and stress reactivity. *The Journal of Clinical Psychiatry*, 74(8), 16662. <https://doi.org/10.4088/JCP.12m08083>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-59. <https://doi.org/10.21427/D7CF7R>
- Hřebíčková, M., & Urbánek, T. (2001). *NEO pětifaktorový osobnostní inventář* (podle NEO Five-Factor Inventory P. T. Costy a R. R. McCrae). Testcentrum 2001.
- Huang, Y., Lee, J., Chen, Z., Perry, M., Cheung, J. H., & Wang, M. (2017). An item-response theory approach to safety climate measurement: The liberty mutual safety climate short scales. *Accident Analysis & Prevention*, 103, 96-104. <https://doi.org/10.1016/j.aap.2017.03.015>
- Ito, M., Oe, Y., Kato, N., Nakajima, S., Fujisato, H., Miyamae, M., Kanie, A., Horikoshi, M., & Norman, S. B. (2015). Validity and clinical interpretability of Overall Anxiety Severity and Impairment Scale (OASIS). *Journal of Affective Disorders*, 170, 217-224. <https://doi.org/10.1016/j.jad.2014.08.045>
- Jackson, D. L., Gillaspay, J. A. Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23. <https://doi.org/10.1037/a0014694>
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19. <https://doi.org/10.1037/0022-006X.59.1.12>
- John, O. P., & Srivastava, S. (1999). *The big five trait taxonomy: History, measurement, and theoretical perspectives*. Guilford Press.
- Kim, J., & Wilson, M. (2020). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement*, 80(4), 726-755. <https://doi.org/10.1177/0013164419892667>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16(9), 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, 64(2), 258-266. <https://doi.org/10.1097/00006842-200203000-00008>
- Lewis, C. A., Shevlin, M. E., Smékal, V., & Dorahy, M. J. (1999). Factor structure and reliability of a Czech translation of the Satisfaction with Life Scale among Czech university students. *Studia psychologica*. Retrieved from

- <https://is.muni.cz/publication/225013/cs/Factor-structure-and-reliability-of-a-Czech-translation-of-the-Satisfaction-With-Life-Scale-among-Czech-university-students/Alan-Lewis-E-Shevlin-Smekal-J-Dorahy>
- Maliňáková, K., Černá, A., Fürstová, J., Čermák, I., Trnka, R., & Tavel, P. (2019). Psychometric analysis of the Guilt and Shame Experience Scale (GSES). *Československá psychologie*, 63(2), 177-192. Retrieved from <http://cejsh.icm.edu.pl/cejsh/element/bw-metal.element.65ac7f69-9a29-482e-996d-5445900da1f0>
- Marsella, A. (2003). Cultural aspects of depressive experience and disorders. *Online Readings in Psychology and Culture*, 10(2). <https://doi.org/10.9707/2307-0919.1081>
- Mikoska, P., Novak, L., Pilarik, L., Bok, T., Fülep, M., & Korinek, R. (2022). *A Czech version of the Overall Anxiety Severity and Impairment Scale (OASIS): Standardization and psychometric properties*. PsyArXiv. Retrieved from psyarxiv.com/s6hab
- Mira, A., González-Robles, A., Molinari, G., Miguel, C., Díaz-García, A., Bretón-López, J., García-Palacios, A., Quero, S., Baños, R., & Botella, C. (2019). Capturing the severity and impairment associated with depression: The overall depression severity and impairment scale (ODSIS) validation in a Spanish clinical sample. *Frontiers in Psychiatry*, 10. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsy.2019.00180>
- Neff, K. D. (2003). The Development and Validation of a Scale to Measure Self-Compassion. *Self and Identity*, 2(3), 223-250. <https://doi.org/10.1080/15298860309027>
- Norman, S. B., Hami Cissell, S., Means-Christensen, A. J., & Stein, M. B. (2006). Development and validation of an Overall Anxiety Severity and Impairment Scale (OASIS). *Depression and Anxiety*, 23(4), 245-249. <https://doi.org/10.1002/da.20182>
- Novak, L. (2022). *Psychtoolbox: Tools for psychology research and psychometrics*. Olomouc University Social Health Institute, Palacky University in Olomouc. Retrieved from <https://gitlab.com/lukas.novak/psychtoolbox>
- Osma, J., Quilez-Orden, A., Suso-Ribera, C., Peris-Baquero, O., Norman, S. B., Bentley, K. H., & Sauer-Zavala, S. (2019). Psychometric properties and validation of the Spanish versions of the overall anxiety and depression severity and impairment scales. *Journal of Affective Disorders*, 252, 9-18. <https://doi.org/10.1016/j.jad.2019.03.063>
- Peters, G.-J. (2021). *Ufs: Quantitative analysis made accessible*. Retrieved from <https://CRAN.R-project.org/package=ufs>
- Prikner, O. (2021). *Vybrané psychometrické charakteristiky škály GAD-7*. 85. Retrieved from <https://is.muni.cz/th/imwek/Diplomova-prace-finis.pdf>
- Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. Retrieved from <https://www.jstatsoft.org/v48/i02/>
- Sandora, J., Novak, L., Brnka, R., Dijk, J. P. van, Tavel, P., & Malinakova, K. (2021). The abbreviated overall anxiety severity and impairment scale (OASIS) and the abbreviated overall depression severity and impairment scale (ODSIS): Psychometric properties and evaluation of the Czech versions. *International Journal of Environmental Research and Public Health*, 18(19), 10337. <https://doi.org/10.3390/ijerph181910337>
- Schermelleh Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23-74.
- Seryjová Juhová, D., Řiháček, T., Cígler, H., Dubovská, E., Saic, M., Černý, M., ... Evans, C. (2018). Česká adaptace dotazníku CORE-OM: vybrané psychometrické charakteristiky. *Československá psychologie*, 62(1), 59-74. Retrieved from <https://www.muni.cz/en/research/publications/1415916>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Tiller, J. W. (2013). Depression and anxiety. *The Medical Journal of Australia*, 199(6), S28-S31.
- Torre, J. A. la, Vilagut, G., Ronaldson, A., Serrano-Blanco, A., Martín, V., Peters, M., Valderas, J., M., Dregan, A., & Alonso, J. (2021). Prevalence and variability of current depressive disorder in 27 European countries: A population-based study. *The Lancet Public Health*, 6(10), e729-e738. [https://doi.org/10.1016/S2468-2667\(21\)00047-5](https://doi.org/10.1016/S2468-2667(21)00047-5)
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70. <https://doi.org/10.1177/109442810031002>

WHO (2016). *Process of translation and adaptation of instruments*. Retrieved from https://www.who.int/substance_abuse/research_tools/translation/en/

Wiger, D. E., & Solberg, K. B. (2001). *Tracking mental health outcomes: A therapist's guide to measuring client progress, analyzing data, and improving your practice*. Wiley.

Winkler, P., Formanek, T., Mlada, K., Kagstrom, A., Mohrova, Z., Mohr, P., & Csemy, L. (2020). Increase in prevalence of current mental disorders in the context of COVID-19: Analysis of repeated nationwide cross-sectional surveys. *Epidemiology and Psychiatric*

Sciences, 29, e173. <https://doi.org/10.1017/S2045796020000888>

Winkler, P., Mohrova, Z., Mlada, K., Kuklova, M., Kagstrom, A., Mohr, P., & Formanek, T. (2021). Prevalence of current mental disorders before and during the second wave of COVID-19 pandemic: An analysis of repeated nationwide cross-sectional surveys. *Journal of Psychiatric Research*, 139, 167-171. <https://doi.org/10.1016/j.jpsychires.2021.05.032>

Zung, W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12(1), 63-70. <https://doi.org/10.1001/archpsyc.1965.01720310065008>

APPENDIX

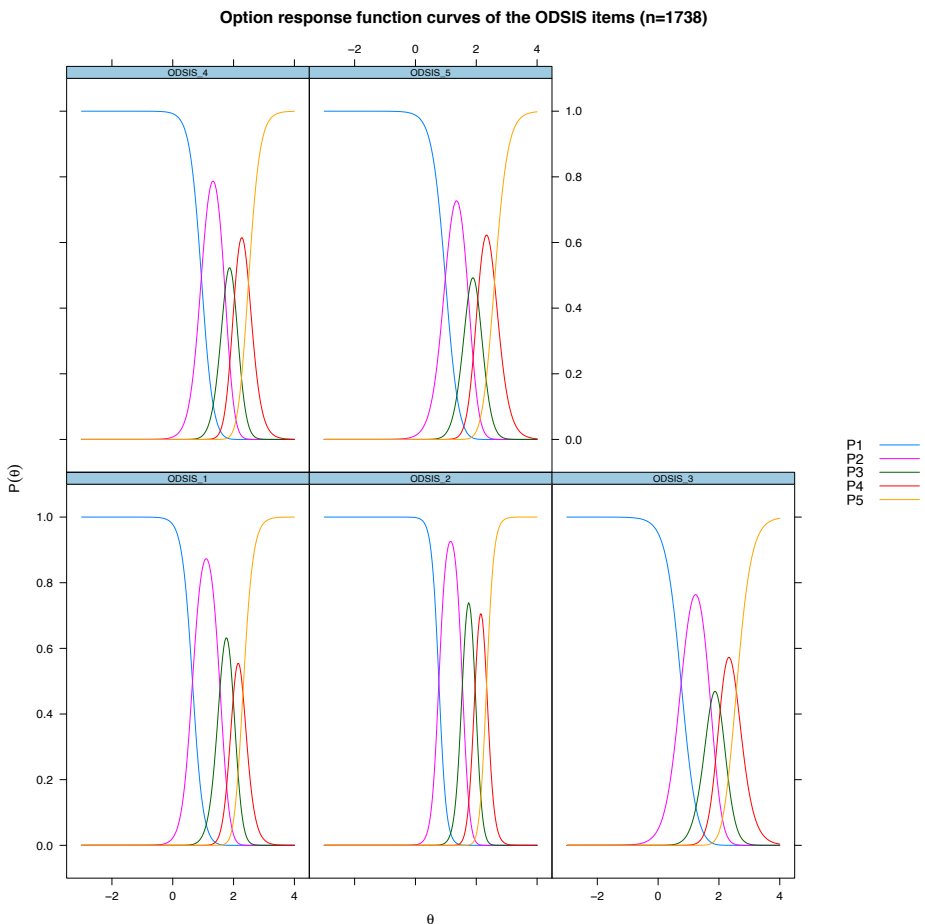


Figure 4 Option response function curves of the ODSIS items

Note. Values on the x-axis refers to level of depression (as a latent variable – θ) standardized to Z-score. The y-axis refers to the probability of endorsing one of the five possible response options – $P(\theta)$. Coloured lines (P1–P5) represent these options sorted from lowest (P1) to highest (P5).

Table 5 presents the frequency of responses to the individual ODSIS items. This table also presents the exact values of difficulty and discrimination parameters.

Table 5 Response frequencies and parameters of the ODSIS items (n = 1738)

Parameters/response options	ODSIS_1	ODSIS_2	ODSIS_3	ODSIS_4	ODSIS_5
α_i	5.806	8.346	3.899	5.188	4.480
b1	0.649	0.771	0.759	0.941	0.983
b2	1.554	1.545	1.728	1.718	1.740
b3	1.987	1.961	2.041	2.031	2.050
b4	2.308	2.339	2.571	2.489	2.599
0/None	71.69%	75.32%	73.88%	79.4%	80.09%
1/Mild	21.23%	18.12%	20.31%	15.54%	14.84%
2/Moderate	5.06%	4.6%	4.09%	3.45%	3.51%
3/Severe	1.55%	1.61%	1.5%	1.38%	1.38%
4/Extreme	0.46%	0.35%	0.23%	0.23%	0.17%

Note. The first column represents item parameters and also response frequencies; ODSIS = Overall Depression Severity and Impairment Scale, α = discrimination parameter, b = difficulty parameter