Journal of Psychosomatic Research 84 (2016) 22–30

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

The construct validity of the Perceived Stress Scale

Marie Germund Nielsen [a,*], Eva Ørnbøl [b], Mogens Vestergaard [a,c], Per Bech [d], Finn Breinholt Larsen [e], Mathias Lasgaard [e,f], Kaj Sparle Christensen [a]

[a] Research Unit for General Practice, Department of Public Health, Aarhus University, Denmark
[b] Research Clinic for Functional Disorders and Psychosomatics, Aarhus University Hospital, Denmark
[c] Section for General Medical Practice, Department of Public Health, Aarhus University, Denmark
[d] Psychiatric Research Unit, Psychiatric Centre North Zealand, Copenhagen University Hospital, Denmark
[e] Public Health and Quality Improvement, Central Denmark Region, Aarhus, Denmark
[f] Department of Psychology, Southern University of Denmark, Odense, Denmark

ARTICLE INFO

Article history:
Received 20 July 2015
Received in revised form 11 March 2016
Accepted 14 March 2016

Keywords:
Construct validity
Perceived Stress Scale
Rasch model
Stress
Dimensionality

ABSTRACT

Objective: Stress impacts the quality of life and is associated with increased risk of mental and physical disorders. The Perceived Stress Scale (PSS) is widely used for measuring psychological distress. Although the instrument was originally defined as a single construct, several studies based on classical test theory suggest that a two-dimensional structure is more dominant. We aimed to explore the construct validity and dimensionality of the PSS-10 using modern test theory to determine if the scale is predominantly for a one- or a two-dimensional model.

Methods: The study population consisted of 32,374 citizens who completed the PSS-10 as part of the Danish National Health Survey in 2010. We investigated the construct validity of the PSS-10 by CFA. We examined the scalability by investigating the fit of the data distribution in a unidimensional Rasch model and performing modification of response categories, persons and items. The scale dimensionality was additionally assessed by Mokken and Rasch analysis.

Results: The PSS-10 did not fit the Rasch model. Item four indicated the largest misfit, and items four and seven displayed disordered thresholds. Unidimensionality could not be established although the data showed improved fit to the Rasch model for the two dimensions respectively with the positive and negative items. Mokken analysis revealed fit to the unidimensional model, but disordered thresholds were shown for item four.

Conclusion: Our large population-based study indicated scalability problems in the current version of the PSS-10. The conducted analysis overall revealed better statistical fit for a two-dimensional than a unidimensional model.

1. Introduction

Stress considerably impacts the quality of life worldwide [1] and is associated with a range of adverse health outcomes, including increased risk of cardiovascular events [2,3], metabolic syndromes [4–6] and mortality [7–9]. Stress can also lead to mental illness which is a burden for the individual but may also cause serious productivity losses with societal implications [10]. Although a stress condition carries a substantial burden, it is merely considered a 'risk factor'. For instance, no diagnosis code for stress exists in the 10th version of the International Classification of Disease (ICD-10) and in the 5th version of the Diagnostic and Statistical Manual of Mental Disorders (DSM V) [11,12].

The Perceived Stress Scale (PSS) is a widely used instrument for measuring stress [13]. The PSS evaluates the degree to which an individual has perceived life as unpredictable, uncontrollable and overloading during the previous month. The PSS also assesses the degree to which external demands seem to exceed the individual's perceived ability to cope [13,14].

The original 14-item scale (PSS-14) was developed in 1983 by S. Cohen et al. [13,15,16], but this first version was later revised and reduced into 10-item and 4-item versions [14]. The PSS-10 was originally defined as a single construct because the 'distinction between the two different dimensions in terms of the positively and negatively scored items, was considered irrelevant' [16]. But exploratory factor analysis (EFA) later indicated that a two-dimensional structure was more dominant in the PSS-10 [17]. A confirmatory factor analysis (CFA) by Andreou et al. confirmed that the one-dimensional model did not provide acceptable fit, while the two-dimensional model tented to show a better fit both in the PSS-10 and PSS-14 [18]. A principal component analysis (PCA) supported the existence of two dimensions: one dimension related to perceived stress (measured by six negatively worded items), while another related to coping ability and stress resilience counter-stress (measured by four positively worded items) [17]. A Turkish study by Örücü and Demir found gender differences in a

* Corresponding author at: Research Unit for General Practice, Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Aarhus C, Denmark.
E-mail address: mm@ph.au.dk (M.G. Nielsen).

http://dx.doi.org/10.1016/j.jpsychores.2016.03.009
0022-3999/

translated version of the PSS-10 [19], whereas a study by Barbosa-Leiker et al. indicated that stress and counter-stress were measured equivalently in men and women by the PSS-10 [20]. Furthermore, a study by Gitchel et al. found that women reported higher levels of perceived stress overall and on the positively worded items, but not on the negatively worded items. The study suggested that gender-related item directionality on the PSS-10 might be the primary biasing factor [21].

The psychometric properties of the different versions of the PSS have been extensively studied in many countries by classical test theory (CTT). Several studies conducted in the general population in a variety of countries have found that Cronbach's α for the total scale ranges between 0.75 and 0.91 [13,18,22,23]. The criterion validity was evaluated by Mitchell et al., who found that the PSS was significantly negatively correlated with the mental component of the Short-Form Health Survey (SF-36) from the Medical Outcome Study (MOS); ($p < 0.05$) and ($r = -0.70$) [24].

Overall the PSS-10 seems to have some unsolved issues as several of the assessment methods (including modern test theories) have indicated problems [17,18].

The present study adopts a broader perspective by further investigating the construct validity and the dimensionality of the PSS-10 by CFA and further by applying modern test theory within the framework of a parametric Rasch analysis and a non-parametric Mokken scale analysis. In addition, we aim to investigate the fit of the Rasch model to PSS-10 data collected from the Danish National Health Survey (DNHS) for a unidimensional model and the two suggested dimensions of the scale.

2. Methods

2.1. Study population

The PSS-10 formed part of a battery of self-report questionnaires on physical and mental health in the DNHS in 2010. The DNHS was based on six random subsamples; one from each of the five Danish regions (mutually exclusive) and a national sample. In this study, we used the population-based sample of 52,400 persons from the Central Denmark Region [25].

All randomly selected individuals received an introductory letter, which briefly described the purpose of the voluntary survey and invited the recipient to complete and return an enclosed paper questionnaire. Data were collected from February to April 2010, and non-respondents received up to three postal reminders within six weeks [25]. In total, 34,168 (65.2%) completed and returned the questionnaire.

Collected data were split into two random samples: a development dataset and a validation dataset. The development dataset was created to modify the PSS to achieve better fit with the Rasch model, while the validation dataset was created to test the modified version of the PSS. To assess the influence of large sample size, we also examined ten subsamples (each of 500) randomly extracted from the development dataset.

2.2. Perceived Stress Scale

All 10 items were rated on a five-point response scale (0 = never, 1 = almost never, 2 = sometimes, 3 = fairly often, 4 = very often). According to the CTT scoring procedure, the responses of the positively stated items (i.e. items 4, 5, 7 and 8) must be reversed, and all item scores must be combined to produce a total score in the range 0–40. A high score indicates a high degree of perceived stress, and no cut-offs were predefined [13].

2.3. Response structure

The response structure for both the Rasch model and the Mokken model reflects a probabilistic Guttman pattern [26]. This implies that, for each responding person, the probability of endorsing an 'easy' item must be higher than the probability of endorsing a more 'difficult item' for each responding person, and the other way around for a ´severe´ item [27].

The category probability curve (CPC) of the Rasch analysis illustrates whether each of the response categories reflects progress in a logical order and whether each response option has the highest probability of occurrence within a specific interval along the logit scale [28,29]. Mokken proposed the use of scalability coefficients in 1971 as a method for describing the influence of Guttman errors on the scale strength. Scalability coefficients based on Mokken scale analysis are specifically used to describe deviations in the obtained data appearing in a perfect Guttman pattern.

2.4. Item response theory

Item response theory (IRT) represents a group of several distinct models which are all based on the assumption that the response to any particular item is a function of the difference between the estimated ability of the person (e.g. the level of stress) and the characteristics of the item. In the Rasch model, this represents the difficulty of the item (e.g. the level of depression implied by the item) [30,31].

2.5. Rasch analysis

Rasch models are parametric latent trait models under the IRT paradigm. Rasch analysis is the formal testing of an outcome scale against a mathematical measurement model and was developed by the Danish mathematician Georg Rasch [32]. A scale that fits the Rasch model provides estimates of a person's level of a latent variable (e.g. stress on an interval scale), which allows the application of subsequent parametric statistics. The model is thus a robust model for measurement of latent traits and should address some of the weaknesses of CTT [33,34].

According to the Rasch model, the probability that a person will affirm an item is a logistic function of the difference between the person's level of, for example, stress and the level of stress expressed by the item [28]. The unidimensional measurement model reflects the basic criterion of invariance, which implies that the instrument is required to work the same way for all subgroups of individuals [35].

Two main distinctions have been made of the Rasch model; a dichotomous version and a polytomous version depending on the number of response categories for each item [35]. The polytomous version can take the form of either the Rating Scale Model introduced by Andrich [36] or the Partial Credit Model introduced by Masters [37]. The Partial Credit Model allows for different distance thresholds for items, whereas the Rating Scale Model assumes the distance between the thresholds to be equal across items.

Rasch suggested that chi-square statistics consisting of both global and local tests (all based on item misfit) be applied for evaluating the fit of data to the model. However, large sample sizes are needed to ensure sufficient power calculations, but large sample sizes generally pose a problem for significance tests based on chi-square statistics, and small differences are often reported to indicate a statistically significant misfit between the data and the model [38].

2.6. Mokken scale analysis

Mokken analysis is a nonparametric model employed to study the properties of a set of items in the IRT framework [39]. The purpose of the Mokken model is to validate an ordinal measure of a latent variable. If a given group of items satisfies the criteria of the Mokken model based on the scalability coefficients, the sum of the responses across items can be used to rank respondents on the latent trait [40]. The assumptions behind the monotone homogeneity model of Mokken (MHMM) are thus unidimensionality, monotonicity and local independence. The assumptions behind the double monotone homogeneity model of