Psychiatry Research 200 (2012) 819–826

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Why the factorial structure of the SCL-90-R is unstable: Comparing patient groups with different levels of psychological distress using Mokken Scale Analysis

Muirne C.S. Paap [a,b,*], Rob R. Meijer [c], Peggy T. Cohen-Kettenis [d], Hertha Richter-Appelt [e], Griet de Cuypere [f], Baudewijntje P.C. Kreukels [d], Geir Pedersen [g], Sigmund Karterud [b], Ulrik F. Malt [a], Ira R. Haraldsen [a,*]

[a] Department of Neuropsychiatry and Psychosomatic Medicine, Oslo University Hospital, Norway
[b] Institute of Clinical Medicine, University of Oslo, Norway
[c] Department of Psychometrics and Statistical Techniques, Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands
[d] Department of Medical Psychology, VU University Medical Center, Amsterdam, The Netherlands
[e] Institute for Sex Research and Forensic Psychiatry, University Hospital Hamburg-Eppendorf, Hamburg, Germany
[f] Department of Sexology and Gender Problems, University Hospital Gent, Belgium
[g] Department for Personality Psychiatry, Clinic for Mental Health and Addiction, Oslo University Hospital, Norway

ARTICLE INFO

Article history:
Received 3 August 2011
Received in revised form
7 March 2012
Accepted 9 March 2012

Keywords:
Item response theory
Validity
Personality disorder
Questionnaire evaluation
Gender identity disorder
Depression

ABSTRACT

Since its introduction, there has been a debate about the validity of the factorial structure of the SCL-90-R. In this study we investigate whether the lack of agreement with respect to the dimensionality can be partly explained by important variables that might differ between samples such as level of psychological distress, the variance of the SCL-90-R scores and sex. Three samples were included: a sample of severely psychiatrically disturbed patients ($n=3078$), a sample of persons with Gender Incongruence (GI; $n=410$) and a sample of depressed patients ($n=223$). A unidimensional pattern of findings were found for the GI sample. For the severely disturbed and depressed sample, a multidimensional pattern was found. In the depressed sample sex differences were found in dimensionality: we found a unidimensional pattern for the females, and a multidimensional one for the males. Our analyses suggest that previously reported conflicting findings with regard to the dimensional structure of the SCL-90-R may be due to at least two factors: (a) level of self-reported distress, and (b) sex. Subscale scores should be used with care in patient groups with low self-reported level of distress.

1. Introduction

The Symptom Checklist-90-Revised (SCL-90-R) (Derogatis, 1994) was designed to cover nine different dimensions of psychological distress; the mean item score across all 90 items with theoretical values ranging from 0 through 4 is referred to as the Global Severity Index (GSI), which is widely used as a global index for psychological distress. Since the introduction of the SCL-90 (-R), there has been a debate about the validity of the factorial structure, which was aptly expressed in the title of the paper 'Factor structure of the SCL-90-R: is there one?' (Cyr et al., 1985). More than two decades have passed since the publication of that paper; however, the debate has still not abated, as recent publications have demonstrated (Olsen et al., 2004; Arrindell et al., 2006; Elliott et al., 2006; Hafkenscheid et al., 2007). On one hand, there is a group of researchers that firmly believe in the multidimensionality of the instrument (Arrindell et al., 2004b, 2004a, 2006), whereas another group has pointed out that alternative models with only one or at most a few factors show an equally good or better fit (Hafkenscheid, 2004; Hafkenscheid et al., 2007). In a recent paper, Paap et al. (2011b) proposed a new scale solution of 7 scales based on a study involving patients referred for a personality disorder (PD); scales were built on two start items that reflected the content of the disorder that corresponded with the specific scale. The new solution included 60 of the 90 items clustered in seven scales: Depression, Agoraphobia, Physical Complaints, Obsessive-Compulsive, Hostility (unchanged), Distrust and Psychoticism. The authors found that most of the new scales discriminated reliably between patients with moderately low scores to moderately high scores.

* Correspondence to: Clinic for Gender Identity Disorder, Department of Neuropsychiatry and Psychosomatic Medicine, Rikshospitalet, NO-0027 Oslo, Norway. Tel.: +47 23074160; fax: +47 23074170.
E-mail addresses: muirne@nxdomain.nl (M.C.S. Paap), i.h.haraldsen@medisin.uio.no (I.R. Haraldsen).

0165-1781/$ - see front matter © 2012 Elsevier Ireland Ltd All rights reserved.
http://dx.doi.org/10.1016/j.psychres.2012.03.012

The items forming the GSI showed low scalability, and the authors concluded that their research findings lent support for a multidimensional model of the SCL-90-R. The authors speculated that the lack of agreement between studies might be due to several factors, such as difference in variance, the existence of structure generating factors, differences in the interpretation of the fit indices, and, finally, the chosen analytic strategy (Paap et al., 2011b).

In the current study, we investigate whether the findings in the study by Paap et al. can be generalized to other patient groups by comparing the dimensionality of the PD sample to that of a sample of persons with Gender Incongruence (GI) and a sample of depressed outpatients. The term 'GI' signifies the incongruence between one's gender identity on one hand, and one's assigned gender and/or one's congenital primary and secondary sex characteristics on the other hand (Kreukels et al., 2010; Meyer-Bahlburg, 2010).[1] Following Kreukels et al., we use GI when referring to patients who have not yet been diagnosed with GID (APA, 1994) or transsexualism (WHO, 1992). We expect the reported level of psychological distress (estimated by the GSI) to be lower in the GI sample than in the depressed sample and PD sample. Haraldsen and Dahl (2000) showed that patients diagnosed with GID had slightly elevated GSI scores when compared to healthy adults, but did not reach the value of 1.0 which is the cut-off for clinically significant symptoms ($GSI_{GID}=0.6$, $GSI_{controls}=0.4$). In contrast, depressed outpatients have been found to exceed the cut-off ($GSI_{DEP}=1.4$) (Leinonen and Niemi, 2007), and so have the patients in the PD sample used in the study by Paap et al. ($GSI_{PD}=1.5$). Our main research questions are:

(1) Is the dimensionality of the SCL-90-R similar for patient groups that differ in level of reported psychological distress?
(2) Are the different factorial solutions found in the literature due to a difference in variance in reported psychological distress?

Following Paap et al. (2011b) and Meijer et al. (2011), Mokken Scale Analysis (MSA; Mokken, 1971) was used to analyze the data. MSA is a nonparametric Item Response Theory (IRT) approach that can be used to explore and test hypotheses about the dimensionality of a data-set, while at the same time resulting in scales adhering to a measurement model.

2. Methods

2.1. Participants

2.1.1. Personality disorder sample: PD_{low} and PD_{high}

This sample consisted of 3078 patients admitted to 14 different day hospitals participating in the Norwegian Network of Personality-Focused Treatment Programs (Karterud et al., 1998), treated in the period from January 1993 through July 2007. This sample was also used in the study by Paap et al. (2011b). Sex ratio and age are depicted in Table 1. Seventy-nine percent were diagnosed with at least one personality disorder (PD). Of the PDs, Avoidant PD was most common (39%), followed by Borderline PD (24%). Extensive comorbidity was common in this group. All patients had at least one axis I disorder. The majority of the patients fulfilled criteria for either Major Depressive Disorder or Dysthymic Disorder (69%), and almost half of the patients were phobic (45% fulfilled criteria for at least one of the following: Agoraphobia, Social Phobia or Specific Phobia). We refer to Paap et al. (2011b) and Karterud et al. (2003) for sociodemographic and diagnostic details. Patients admitted before 1996 were diagnosed according to the DSM-III-R (APA, 1987) and patients admitted from 1996 onwards according to the DSM-IV (APA, 1994).

In the study by Paap et al. (2011b), two subgroups were created based on clinical criteria: the first group existed of patients with a clinical disorder (CD) only (GSI=1.3), and the second group of patients diagnosed with a PD in addition to a CD (GSI=1.6). Since the focus of the current study is on the impact of psychological distress on dimensionality, we chose to use a different criterion to create two subgroups in the current study. To maximize the difference in GSI scores in the resulting subgroups while at the same time create subgroups that showed similar variance of GSI scores as the GI and depression samples, the total group of 3078 patients was divided along the median GSI-score (1.53). The group consisting of patients with a GSI-score through 1.53 are referred to as the PD_{low} group ($n=1528$, mean age$=35\pm9$ years) and the group of patients with a GSI-score of 1.53 or higher as the PD_{high} group ($n=1550$, mean age$=35\pm9$ years).

All participating hospitals complied with the diagnostic and data collection procedures required for membership in the Norwegian Network. All data registered by the different hospitals were collected regularly in a central, anonymised database, administrated by the Department of Personality Psychiatry, Oslo University Hospital. All patients gave written consent and the procedures were approved by the State Data Inspectorate and the Regional Committee for Medical Research and Ethics.

2.1.2. Gender incongruence sample

This sample consisted of 410 persons referred to four Gender Identity Disorder (GID) clinics: Ghent (Belgium), Hamburg (Germany), Amsterdam (the Netherlands) and Oslo (Norway). The data collection took place within the framework of the 'European Network for the Investigation of Gender Incongruence' (ENIGI) initiative (Kreukels et al., 2010). This network was created in order to improve comparability of data pertaining to gender incongruence (GI) and GID across clinics, as well as diagnostic transparency (Paap et al., 2011a). The ENIGI study includes applicants that were seen at GID clinics in Ghent, Hamburg, Amsterdam, and Oslo from the start of January 2007. In the current study all new applicants that were seen between January 2007 and December 2009 and whose data had been entered in the database, were at least 16 years of age at their first visit, and who had filled out the SCL-90-R were included. Sex ratio (reported sex corresponds to natal sex) and age are depicted in Table 1. At the time of data analysis, 56% of the total sample had been diagnosed with GID, 10% with another disorder pertaining to gender incongruent feelings (such as transvestic fetishism or GID NOS) and the remaining 34% had not yet received a diagnosis. The four participating clinics complied with the diagnostic and data collection procedures required for membership in the ENIGI initiative. All data registered by the different clinics were collected regularly in a central, anonymised database, administrated at the Oslo University Hospital. All patients gave written consent and the procedures were approved by the regional ethical committees.

2.1.3. Depression sample

This sample consisted of 223 patients who had been referred to the Department of Neuropsychiatry and Psychosomatic Medicine at Oslo University Hospital and fulfilled the DSM-IV (APA, 1994) criteria for Major Depressive Disorder or Dysthymic Disorder. The patients were at least 18 years old at the first visit, and were seen between January 2005 and December 2008. Sex ratio and age are depicted in Table 1. Seventy-four percent of the patients fulfilled criteria for at least one other axis I disorder, of which a phobic disorder was most common (46% fulfilled criteria for either Agoraphobia, Social Phobia or Specific Phobia), followed by Generalised Anxiety Disorder (37%). The M.I.N.I. (Sheehan and Lecrubier, 1994) was used to screen for axis I disorders. All patients gave written consent and the procedures were approved by the State Data Inspectorate and the Regional Committee for Medical Research and Ethics.

2.2. Measures

All patients completed a number of self-report measures prior to or directly after one of the first consultations, including the Symptom Checklist 90-Revised (SCL-90-R: Derogatis, 1994). The instrument was designed to measure nine symptom dimensions (comprising a total of 83 items): somatization (Som), interpersonal sensitivity (Int), depression (Dep), anxiety (Anx), phobic anxiety (Pho), obsession-compulsion (Obs), hostility (Hos), paranoid ideation (Par), and psychoticism (Psy), and includes 7 additional items. Each item is scored on a scale ranging from 0 ('not at all') through 4 ('extremely'). The mean score on all 90 items (including the 7 additional items) is referred to as the Global Severity Index (GSI; range 0–4) and is widely used as a global index for psychological distress.

2.3. Investigating dimensionality: Mokken Scale Analysis (MSA)

To investigate the dimensionality of the SCL-90-R, MSA was used (Mokken, 1971, 1997). MSA is a nonparametric Item Response Theory (IRT) approach that can be used to explore and test hypotheses about the dimensionality of a data-set. MSA can be used in a confirmatory or exploratory way. In either case, it assesses whether the

[1] Note that the DSM5 Work Group for Sexual and Gender Identity Disorders first considered proposing the replacement of "GID" with the term "Gender Incongruence", but that the current recommendation is replacing "GID" with "Gender Dysphoria". However, gender incongruent feelings are still considered core to the phenomenon of GID/Gender Dysphoria. In this paper, we use the term "GI" to signify people who experience gender incongruent feelings, but who have not yet been diagnosed.