The Dutch Symptom Checklist-90-Revised

Smits, Iris A. M.; Timmerman, Marieke E.; Barelds, Dick P. H.; Meijer, Rob R.

Published in:
European Journal of Psychological Assessment

DOI:
10.1027/1015-5759/a000233

Publication date:
2015

Original Article

The Dutch Symptom Checklist-90-Revised

Is the Use of the Subscales Justified?

Iris A. M. Smits, Marieke E. Timmerman, Dick P. H. Barelds, and Rob R. Meijer

University of Groningen, Faculty of Behavioural and Social Sciences, The Netherlands

Abstract. The Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1977, 1994) was constructed to measure both general psychological distress and specific primary symptoms of distress. In this study, we evaluated to what extent the scale scores of the Dutch SCL-90-R reflect general and/or specific aspects of psychological distress in a psychiatric outpatients sample (N = 1,842), using a hierarchical factor model. The results revealed that the total scale score measures general psychological distress, with high reliability. The subscale scores Sleep Difficulties, Agoraphobia, Hostility, and Somatization reflect the specific primary symptoms reasonably well, with high reliability. The subscale score Depression hardly measures specific symptoms of distress, but instead a very common construct as is measured with the total scale of the SCL-90-R. The use of the Depression subscale score beyond the total scale score of the SCL-90-R appears therefore of limited value in clinical practice.

Keywords: SCL-90-R, psychological distress, hierarchical model, bifactor model

The Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1977, 1994) is one of the most frequently used scales in clinical practice (e.g., Evers et al., 2012). The SCL-90-R is used to diagnose, monitor, and screen clients on both psychological distress and multiple aspects of psychological distress. The total scale score of the SCL-90-R is used as an indicator for general psychological distress and the subscale scores are used as indicators for specific primary symptoms, such as anxiety, depression, and somatization.

These specific primary symptoms reflect "clusters of primary symptoms" (Derogatis and Cleary, 1977). Herewith, it conforms to the description of a syndrome as indicating a group of symptoms which consistently co-occur. However, it does not necessarily conform to the other description of a syndrome, namely as referring to a specific condition (e.g., as described in *The Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; *DSM–5*; American Psychiatric Association, 2013) that is characterized by a set of associated symptoms. That is, the primary symptoms are neither directly nor completely related to (DSM-5) syndromes. There are, however, primary symptoms that bear a strong relationship to a specific (DSM-5) syndrome. An example is the primary symptom Obsessive Compulsive of the SCL-90-R, which is very closely related to the DSM-5 (American Psychiatric Association, 2013) clinical syndrome Obsessive Compulsive Disorder (see Derogatis & Cleary, 1977). Other primary symptoms, such as Sleep Difficulties, play a part in several (DSM-5) syndromes.

The SCL-90-R subscale scores are thus interpreted in terms of primary symptom dimensions, which are to some extent related to (DSM-5) syndromes. Moreover, in addition to the primary symptom dimensions, the SCL-90-R quantifies psychopathology in terms of overall psychological distress. In particular, the SCL-90-R total scale score is used as a global index of psychological distress and the SCL-90-R subscale scores are used as indices of specific aspects of psychological distress. A crucial question is whether the use of both the total scale score and the various subscale scores of the SCL-90-R is justified. A necessary prerequisite for justifying this multipurpose use is that both the general construct of psychological distress and its specific primary symptoms are theoretically and empirically distinguishable.

The authors of the SCL-90-R and its precursors, the SCL-90 (Derogatis, Lipman, & Covi, 1973) and the HSCL (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974), offered both theoretically and empirically based arguments for the use of both the total scale score and the subscale scores. However, further empirical research, as conducted by many authors, is inconclusive with respect to these claims. Researchers have made different recommendations with respect to the dimensionality of the SCL-90-R. On the one hand, a large number of studies suggested that the SCL-90-R is described best as a unidimensional questionnaire, measuring mainly the general factor of psychological distress (see e.g., Cyr, McKenna-Foley, & Peacock, 1985; Vassend & Skrondal, 1999). These studies question the use of the subscale scores. On the other hand, several studies suggested that the SCL-90-R is described best as a multidimensional inventory (see e.g., Arrindell, Barelds, Janssen, Buwalda, & van der Ende, 2006; Holcomb, Adams, & Ponder, 1983). Thus, previous studies were not

DOI: 10.1027/1015-5759/a000233

univocal with respect to whether the SCL-90-R can be described best as a unidimensional or as a multidimensional questionnaire. More importantly, the implications of those results for the usefulness of the total scale score and the subscale scores of the SCL-90-R are still unclear.

For clinical practice, it is of key importance to know how one should interpret the total score and subscale scores of the SCL-90-R. This interpretation depends on the amount of specific variance explained by the subscale scores and the total scale scores (see e.g., Reise, Bonifay, & Haviland, 2013). For a subscale to be useful in clinical practice, the amount of specific variance should be substantial. If not, the subscale variance would be mainly due to the shared variance with all other subscales and due to measurement error. Both sources of nonspecific variances have detrimental effects on the usefulness of subscale scores. Though this is widely acknowledged with respect to measurement error, the amount of variance shared with all other subscales is typically neglected. If the latter amount would be substantial, the subscale scores measure for a substantial part the same or highly correlated symptoms as the total scale and, as a result, reporting subscale scores adds little to reporting only the total score. In this study, we investigate the degree to which the subscale scores reflect specific aspects of psychological distress apart from general psychological distress. In doing this, we also evaluate the common clinical practice to use both the total and subscale scores of the SCL-90-R.

Investigating the Unique Contribution of Subscales

Investigating the unique contribution of subscale scores above the total score requires a measurement model that takes into account the different levels of the construct hierarchy and considers the importance of the specific factors given the general factor. Hierarchical and higher-order factor models are often used to investigate such structures (e.g., Brunner, Nagy, & Wilhelm, 2012; Reise, 2012; Reise et al., 2013).

Figure 1 shows an illustrative representation of a (a) higher-order model and a (b) hierarchical model for the Dutch SCL-90-R. In a higher-order model a higher-order factor (here indicated as G) contributes to the correlations between the lower-order factors (the specific factors Agoraphobia, Anxiety, etc.). In a hierarchical model all factors are at the same level, but at different layers. The essential difference is that a higher-order model is more constrained than a hierarchical model. Specifically, each higher-order model can be represented as a hierarchical model with a proportionality constraint that enforces each item to have the same ratio of its loadings on the general factor and the specific factor. An advantage of the hierarchical representation is that the specific contribution of the specific factors above the general factor can be inferred directly. Thus, for the analysis of clinical instruments, a hierarchical representation is a useful approach to explore the unique contribution of the subscale scores above the total scale scores. In empirical practice, it is safest to consider the (unconstrained) hierarchical model, as proportionality is by no means guaranteed.

Only a few studies have used a hierarchical model to analyze the SCL-90-R. These studies, however, could not provide a full answer to the question whether the SCL-90-R in its current form can reliably distinguish all specific factors above the general factor. In the study by Vassend and Skrondal (1999), a hierarchical model was used to analyze the SCL-90-R, but in that study the authors did not investigate the unique contribution of the subscale scores above the total scale scores. Hafkenscheid, Maassen, and Veeninga (2007) did investigate the unique contribution of the subscale scores, but they considered a specific factorial structure of the SCL-90-R that deviates from the structure of the SCL-90-R as it is commonly used in clinical practice.

The aim of the present study is to examine whether the SCL-90-R can reliably measure psychological distress as well as multiple aspects of psychological distress. In particular, we concentrate on the question to what extent the observed score variances on the SCL-90-R subscales and total scale can be attributed to the general factor, to one or more specific factors, and to measurement error. Herewith, we want to evaluate the quality of the subscale scores of the SCL-90-R as measures of specific primary symptoms of psychological distress.

Materials and Methods

Participants

We analyzed the data of a group of psychiatric outpatients that was used by Arrindell and Ettema (2003) to construct norms for psychiatric outpatients in the Netherlands. The sample consisted of $N = 1{,}872$ psychiatric outpatients who completed the SCL-90-R during intake at a mental health clinic or university research clinic located in different parts of the Netherlands between 1989 and 2002. Data of individual diagnoses were not available, but given that various mental health organizations in the Netherlands took part at the study and that all outpatients who received an intake were included in the sample, it is probable that the distribution of psychiatric disorders is consistent with the prevalence of psychiatric disorders in the Netherlands. This would imply that the outpatients were seen for a variety of DSM-IV Axis I and Axis II disorders, with the most prevalent diagnoses anxiety and mood disorders (e.g., Vollebergh et al., 2003).

We excluded cases of patients with more than half of the items (i.e., 45 items or more) missing (excluded: 19 cases; 1%). Furthermore, cases of patients missing a major part of the items of a specific subscale score (i.e., less than 2 valid items per subscale score) were excluded (excluded: 11 cases; 0.6%). The resulting total sample consisted of $N = 1{,}842$ psychiatric outpatients with a mean age of

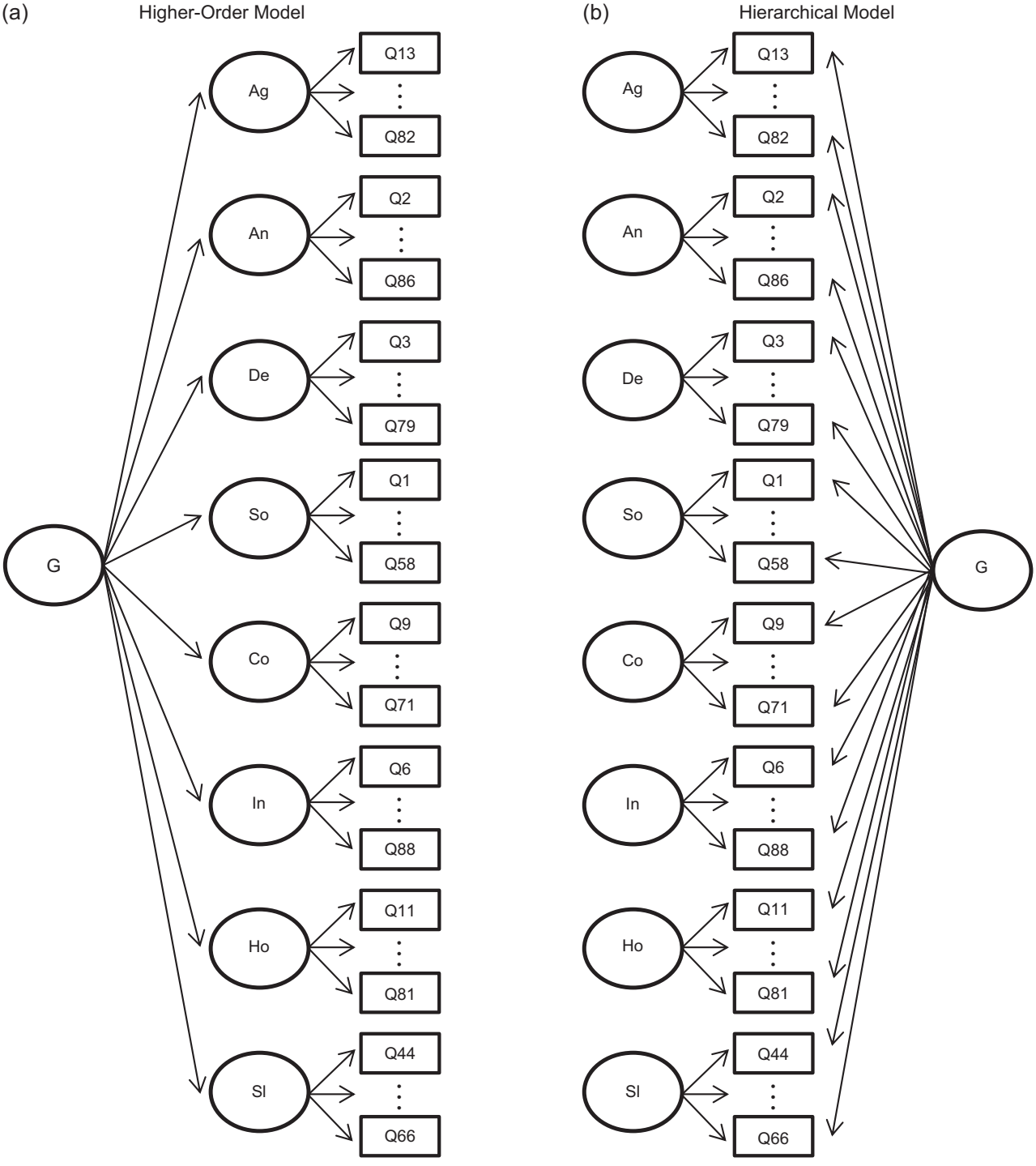

Figure 1. Representation of a (a) Higher Order Model and a (b) Hierarchical Model. Circles indicate factors, with the following abbreviated names (and full names): G (General); Ag (Agoraphobia); An (Anxiety); De (Depression); So (Somatization); Co (Cognitive-Performance Deficits); In (Interpersonal Sensitivity); Ho (Hostility); Sl (Sleep Difficulties). The uncorrelated error terms are not explicitly depicted for ease of representation.

35.2 (*SD* = 11.0), among which 729 males, 1,109 females, and 4 of unknown gender. This dataset contained a limited number of missing data (summary statistics of the percentage of missing data per patient: min = 0.00%, max = 47.78%; 10th, 50th, 90th percentiles = 0.00%, 0.00%, 3.33%, respectively). The missing data were imputed using "Two-Way imputation with normally distributed Errors" (van Ginkel, van der Ark, & Sijtsma, 2007).

To exclude that the predefined factor structure of the data on which the imputation procedure is based would affect the results, we imputed the data twice: (1) on the basis of all available data, considering it as unidimensional data, and (2) on the basis of the available data of the specific subscales, considering it as multidimensional data.

Instrument

The SCL-90-R contains 90 items, each offering a short description of a symptom. Participants rate, on a five-point rating scale, the degree to which they have experienced each of the symptoms during the past week. In the present study, we used the Dutch version of the SCL-90-R by Arrindell and Ettema (1986, 2003). The Dutch structure of the SCL-90-R deviates slightly from the original structure of the SCL-90-R as reported by Derogatis (1977, 1994). Differences between the Dutch and the original structure of the SCL-90-R are relatively small though, with only some items shifting scales (see Table 1 for an overview of the partitioning of the items into scales for both the Dutch and original versions). Arrindell and Ettema (2003) concluded that all eight Dutch dimensions have been described previously in the literature on the structure of the (American) SCL-90-R. The Dutch version of the SCL-90-R consists of the subscales Anxiety, Agoraphobia, Depression, Somatization, Cognitive-Performance Deficits, Interpersonal Sensitivity, Hostility, and Sleep Difficulties. In addition, the SCL-90-R comprises a total scale score for psychological distress, made up of all 90 items (Arrindell & Ettema, 2003). This total score is obtained by summing the scores over the 90 items, and is denoted as psychological distress or psychoneuroticism. This total score has a similar interpretation as the Global Severity Index (GSI), which is obtained by averaging the scores over the 90 items.

Analyses

We analyzed the data of the psychiatric outpatient sample with a two-layer confirmatory hierarchical factor model for ordered polytomous items, where each of the 90 items loads on both its associated specific factor and the general factor. Here, the specific factors represent the individual constructs as measured by the subscale scores of the SCL-90-R, and the general factor represents the construct psychological distress, as measured by the total scale score of the SCL-90-R. The general factor and specific factors are orthogonal. Possible inter-correlations between the subscales are captured through the general factor.

We will now describe the two-layer hierarchical factor model more in detail. If $\mathbf{y}_i^*$ denotes the p-dimensional vector of continuous latent scores of subject i on p items, we specify the following linear factor model for $\mathbf{y}_i^*$:

$$\boldsymbol{y}_i^* = \boldsymbol{\nu} + \boldsymbol{\lambda}_{iG}\,\eta_{iG} + \boldsymbol{\lambda}_{iS1}\eta_{iS1} + \boldsymbol{\lambda}_{iS2}\eta_{iS2} + \ldots + \boldsymbol{\lambda}_{iS8}\eta_{iS8} + \boldsymbol{\varepsilon}_i, \quad (1)$$

where $\boldsymbol{\nu}$ is a p-dimensional vector of intercepts, $\boldsymbol{\lambda}_{iG}$ is a p-dimensional vector of factor loadings of the general factor, $\boldsymbol{\lambda}_{iS1}$ to $\boldsymbol{\lambda}_{iS8}$ are p-dimensional vectors of factor loadings of the specific factors, η_{iG} is the common factor score of subject i on the general factor, η_{iS1} to η_{iS8} are the common factor scores of subject i on the specific factors, and $\boldsymbol{\varepsilon}_i$ is a p-dimensional vector of residuals. It is assumed that $E(\boldsymbol{\varepsilon}_i) = 0$, that $\boldsymbol{\varepsilon}_i$ are mutually independent, and that η_{iG}, $\eta_{iS1},\ldots,\eta_{iS8}$ and $\boldsymbol{\varepsilon}_i$ are independent.

The continuous latent variables $\boldsymbol{y}_i^*$ are related to the observed ordered polytomous variables $\boldsymbol{y}_i$ via thresholds τ as:

$$y_i = c \quad \text{if} \quad \tau_{c-1} < y_i^* < \tau_c, \quad (2)$$

for categories $c = 1, \ldots, C$, with C the number of categories (here: 5).

For identification purposes, the means ($\boldsymbol{\alpha}$) of the common factors (η) are fixed at $\boldsymbol{\alpha} = \mathbf{0}$; the variances of the common factors are fixed at one, and the common factors are mutually uncorrelated; the intercepts ($\boldsymbol{\nu}$) of the latent variables $\mathbf{y}^*$ are fixed at $\boldsymbol{\nu} = \mathbf{0}$, resulting in $\boldsymbol{\mu}^* = \mathbf{0}$, and the variances of the latent variables $\mathbf{y}^*$ are fixed at one $(diag(\boldsymbol{\Sigma}^*) = \mathbf{1})$.

The model was fitted using WLSMV estimation, a suitable procedure for polytomous data, in the Mplus program (Muthén & Muthén, 1998–2007). To determine the model fit, we considered the root-mean-square error of approximation (RMSEA; Brown & Cudeck, 1993; Steiger, 1990) and the comparative fit index (CFI; Bentler, 1990). The prevailing convention is that a RMSEA value of .08 or smaller, and a CFI value of .95 or larger indicates an acceptable fit (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003). For completeness, we also report the chi-square. If the model did not fit adequately, we inspected the modification indices to determine the cause of the misfit. We added the parameters in the model if they resulted in an appreciable improvement in fit and were interpretable with respect to content, while we aimed at following the structure underlying the clinical use of the SCL-90-R as closely as possible. The latter implied only allowing model modifications if it were necessary for an acceptable model fit and refraining from allowing relationships between items and specific factors to which the item is not associated in the current use of the SCL-90-R.

On the basis of the final hierarchical model, we computed for each scale the reliability as the proportion of variance of the scale involved accounted for by all factors (ω_k; see e.g., Brunner & Süβ, 2005; Raykov, 2004; Reise et al., 2013; Zinbarg, Revelle, Yovel, & Li, 2005). Furthermore, for the total scale score, we computed the proportion of variance accounted for by the general factor (ω_H), and, for each subscale, the proportion of variance accounted for by the specific factors associated with the scales involved (ω_{N_k}). The latter reflects the proportion of common variance unique from the general factor. In order to roughly classify scales on their unique proportion of variance, we consider this unique variance to be substantial for a value $\omega_{N_k} \geq .30$, moderate for a value $.20 \leq \omega_{N_k} < .30$, and low for a value $\omega_{N_k} < .20$.

Table 1. The final hierarchical factor model. The SCL-90-R Items (with original subscales), with Loadings (and Standard Errors) on the General Psychological Distress Factor (λ_g),General Anxiety Factor (λ_{gAnx}), and Specific Factors (λ_s).

Subscale	λ_g	λ_{gAnx}	λ_s
Agoraphobia			
Q13 (PhobAnx)	**.57*** (.02)	**.40*** (.02)	**.53*** (.02)
Q25 (PhobAnx)	**.56*** (.02)	**.41*** (.03)	**.47*** (.02)
Q47 (PhobAnx)	**.46*** (.03)	**.46*** (.03)	**.58*** (.02)
Q50 (PhobAnx)	**.52*** (.02)	**.46*** (.02)	**.35*** (.02)
Q70 (PhobAnx)	**.54*** (.02)	**.30*** (.02)	**.61*** (.02)
Q75 (PhobAnx)	**.64*** (.02)	.24* (.02)	.16* (.03)
Q82 (PhobAnx)	**.48*** (.03)	**.43*** (.03)	.28* (.03)
Anxiety			
Q02 (Anx)	**.59*** (.02)	**.33*** (.03)	**.50*** (.03)
Q17 (Anx)	**.57*** (.02)	**.34*** (.03)	**.44*** (.03)
Q23 (Anx)	**.63*** (.02)	**.49*** (.02)	−.11* (.03)
Q33 (Anx)	**.62*** (.02)	**.56*** (.02)	−.12* (.03)
Q39 (Anx)	**.57*** (.02)	.29* (.03)	.23* (.03)
Q57 (Anx)	**.72*** (.01)	.14* (.03)	.23* (.02)
Q72 (Anx)	**.60*** (.02)	**.65*** (.02)	−.02[b] (.03)
Q78 (Anx)	**.61*** (.02)	.03[a] (.02)	.04[c] (.03)
Q80 (Anx)	**.66*** (.02)	.28* (.02)	−.23* (.03)
Q86 (Anx)	**.56*** (.02)	**.45*** (.02)	−.25* (.03)
Depression			
Q03 (Dep)	**.65*** (.02)		.20* (.02)
Q05 (Dep)	**.34*** (.02)		.14* (.03)
Q14 (Dep)	**.55*** (.02)		**.32*** (.02)
Q15 (Dep)	**.63*** (.02)		**.34*** (.03)
Q19 (Other)	**.49*** (.02)		.17* (.03)
Q20 (Dep)	**.48*** (.02)		.21* (.03)
Q22 (Dep)	**.67*** (.02)		.16* (.02)
Q26 (Dep)	**.64*** (.02)		.11* (.02)
Q29 (Dep)	**.67*** (.01)		**.36*** (.02)
Q30 (Dep)	**.72*** (.01)		**.55*** (.02)
Q31 (Dep)	**.70*** (.02)		**.35*** (.02)
Q32 (Dep)	**.65*** (.02)		**.42*** (.02)
Q51 (OCD)	**.70*** (.01)		**.39*** (.02)
Q54 (Dep)	**.68*** (.01)		**.33*** (.02)
Q59 (Other)	**.63*** (.02)		.22* (.03)
Q79 (Dep)	**.74*** (.01)		.09* (.02)
Somatization			
Q01 (Som)	**.42*** (.02)		**.35*** (.03)
Q04 (Som)	**.52*** (.02)		**.39*** (.02)
Q12 (Som)	**.51*** (.02)		**.39*** (.02)
Q27 (Som)	**.40*** (.02)		**.42*** (.02)
Q40 (Som)	**.58*** (.02)		**.37*** (.02)
Q42 (Som)	**.48*** (.02)		**.56*** (.02)
Q48 (Som)	**.54*** (.02)		**.38*** (.02)
Q49 (Som)	**.61*** (.02)		**.34*** (.02)
Q52 (Som)	**.55*** (.02)		**.50*** (.02)
Q53 (Som)	**.60*** (.02)		.19* (.02)
Q56 (Som)	**.64*** (.02)		**.38*** (.02)
Q58 (Som)	**.60*** (.02)		**.54*** (.02)
Cognitive-performance deficits			
Q09 (OCD)	**.47*** (.02)		**.55*** (.02)
Q10 (OCD)	**.53*** (.02)		**.40*** (.02)
Q28 (OCD)	**.59*** (.02)		.21* (.02)
Q38 (OCD)	**.54*** (.02)		**.44*** (.02)
Q45 (OCD)	**.57*** (.02)		**.42*** (.02)

(Continued on next page)

Table 1. (Continued)

Subscale	λ_g	λ_{gAnx}	λ_s
Q46 (OCD)	**.58*** (.02)		.26* (.02)
Q55 (OCD)	**.68*** (.01)		**.43*** (.02)
Q65 (OCD)	**.49*** (.03)		.19* (.04)
Q71 (Dep)	**.71*** (.01)		.22* (.02)
Interpersonal sensitivity			
Q06 (IntSens)	**.44*** (.02)		.27* (.02)
Q07 (Psychot)	**.52*** (.02)		**.32*** (.03)
Q08 (ParaIdeat)	**.45*** (.02)		.27* (.02)
Q18 (ParaIdeat)	**.63*** (.02)		**.49*** (.02)
Q21 (IntSens)	**.49*** (.02)		**.59*** (.02)
Q34 (IntSens)	**.67*** (.01)		**.38*** (.02)
Q35 (Psychot)	**.57*** (.02)		**.48*** (.02)
Q36 (IntSens)	**.64*** (.02)		**.34*** (.02)
Q37 (IntSens)	**.61*** (.02)		.18* (.03)
Q41 (IntSens)	**.62*** (.02)		.22* (.03)
Q43 (ParaIdeat)	**.60*** (.02)		**.34*** (.02)
Q61 (IntSens)	**.59*** (.02)		**.53*** (.02)
Q68 (ParaIdeat)	**.53*** (.02)		.27* (.02)
Q69 (IntSens)	**.62*** (.02)		**.45*** (.02)
Q73 (IntSens)	**.58*** (.02)		**.39*** (.02)
Q76 (ParaIdeat)	**.58*** (.02)		.27* (.03)
Q83 (ParaIdeat)	**.61*** (.02)		**.35*** (.03)
Q88 (Psychot)	**.55*** (.02)		.22* (.02)
Hostility			
Q11 (AngHost)	**.64*** (.02)		.24* (.02)
Q24 (AngHost)	**.52*** (.02)		**.68*** (.02)
Q63 (AngHost)	**.45*** (.03)		**.61*** (.02)
Q67 (AngHost)	**.54*** (.02)		**.65*** (.02)
Q74 (AngHost)	**.49*** (.02)		**.49*** (.02)
Q81 (AngHost)	**.50*** (.02)		**.75*** (.02)
Sleep difficulties			
Q44 (Other)	**.57*** (.02)		**.55*** (.02)
Q64 (Other)	**.41*** (.02)		**.61*** (.02)
Q66 (Other)	**.59*** (.02)		**.78*** (.02)
Items not included in any specific factor			
Q16 (Psychot)	**.60*** (.03)		
Q60 (Other)	.23* (.03)		
Q62 (Psychot)	**.59*** (.02)		
Q77 (Psychot)	**.72*** (.01)		
Q84 (Psychot)	**.41*** (.03)		
Q85 (Psychot)	**.53*** (.02)		
Q87 (Psychot)	**.59*** (.02)		
Q89 (Other)	**.60*** (.02)		
Q90 (Psychot)	**.64*** (.02)		

Notes. PhobAnx = Phobic Anxiety; Anx = Anxiety; Dep = Depression; Other = Items not included in any specific factor; OCD=Obsessive-Compulsive; Som = Somatization; IntSens = Interpersonal Sensitivity; Psychot = Psychoticism; ParaIdeat = Paranoid Ideation; AngHost = Anger-Hostility. Loadings $\geq$ .30 in absolute value are indicated in bold face. *$p < .005$; [a]$p = .224$; [b]$p = .482$; [c]$p = .132$.

Results

Measurement Model

We fitted the hierarchical factor model to both the unidimensional and multidimensional imputed data. Because the results of those two data sets appeared to be highly similar (mean absolute deviation in estimated lambda's = .005, *SD* = .005), we only present the results from the multidimensional imputed data.

First, we fitted the two-layer hierarchical model. This model showed moderate fit ($\chi^2(3834) = 23977.921$, $p < .005$; RMSEA = .053, 90% CIs[.053, .054]; CFI = .854). Inspection of the modification indices

Table 2. The proportion of variance explained by all Factors (ω_k), the proportion of variance of the total scale explained by the general factor (ω_H), and the proportion of variance of the subscales explained by the non-general Factors (ω_{N_k})

Scale	ω_k	ω_H	ω_{N_k}
Total scale	.98	.93	
Subscale			
Sleep difficulties	.88		.53
Agoraphobia	.92		.49
Hostility	.91		.49
Somatization	.91		.33
Interpersonal sensitivity	.94		.26
Anxiety	.93		.24
Cognitive-performance deficits	.88		.24
Depression	.93		.15

revealed that the largest modification index pertains to the parameter that reflects the constraint on the correlation between the subscales Anxiety and Agoraphobia. Since these subscales are theoretically strongly related, in the next step, we added a layer to the hierarchical model with one factor that is associated to the items of the subscales Anxiety and Agoraphobia. This three-layer hierarchical model had a reasonable fit ($\chi^2(3817) = 22098.193$, $p < .005$; RMSEA = .051, 90% CIs[.050, .052]; CFI = .868). The RMSEA value indicates an acceptable fit, though the CFI value is somewhat low. We took this model as the final model. The loadings of this model are presented in Table 1.

Implication for Strength of Scales

Total Scale

The reliability of the total scale score (ω_k), which equals the proportion of variance accounted for by all factors, was .98 (see Table 2), which is large. A substantial proportion of variance of the total scale score is accounted for by variation on the general factor (ω_H = .93, see Table 2). This suggests that the 90 items of the total scale indeed measure a common construct, which is generally referred to as psychological distress. Furthermore, it indicates that the different aspects of psychological distress have much in common, which is also reflected in the moderate to high inter-correlations between the subscales (inter-correlations ranged from .301 to .719, see Table 3). When considering the size of the general loadings per subscale, it is apparent that particularly items belonging to the subscale Depression and Anxiety load high on the general factor (see Table 1). It is these scales that also correlate relatively highly with the other subscales (see Table 3). This suggests that the general factor psychological distress is particularly well represented by depression and anxiety symptoms.

Subscales

Table 2 shows that the reliability of the subscales is generally high (with ω_k ranging from .88 to .94). However, the subscales differ much in how large the proportion of variance is that is unique from the general factor. Sleep Difficulties, Agoraphobia, Hostility, and Somatization reflect a substantial proportion of common variance that is unique from the general factor (i.e., $\omega_{N_k} \geq .30$). Interpersonal Sensitivity, Anxiety, and Cognitive-Performance Deficits reflect a moderate proportion of unique variance (i.e., $.20 \leq \omega_{N_k} < .30$). Depression shows a low (i.e., $\omega_{N_k} < .20$) proportion of unique variance.

Based on these findings, we conclude that the subscales Sleep Difficulties, Agoraphobia, Hostility, and Somatization substantially reflect unique primary symptoms of psychological distress apart from general psychological distress. The latter holds for the Sensitivity, Anxiety, and Cognitive-Performance Deficits subscales as well, though to a moderate extent only. The subscale Depression hardly reflects unique primary symptoms, implying that this scale does not measure anything substantial beyond general psychological distress.

Note that for the subscale Anxiety (Table 1), the items that reflect physical anxiety symptoms have generally positive loadings on the specific anxiety factor and the items reflecting mental anxiety symptoms have generally negative loadings on the specific anxiety factor. Those opposite signs in the specific anxiety factor result in a reduction of the proportion of variance unique from the general factor. This cannot be ameliorated by mirroring the items, since

Table 3. Correlations between subscale scores of the SCL-90-R

Subscale	Ag	An	De	So	Co	In	Ho	Sl
Ag	1							
An	**.706**	1						
De	.477	**.685**	1					
So	**.568**	**.714**	**.628**	1				
Co	.490	**.619**	**.697**	**.610**	1			
In	**.521**	**.621**	**.719**	**.532**	**.654**	1		
Ho	.324	.461	**.531**	.433	.461	**.593**	1	
Sl	.301	.478	**.540**	**.512**	.449	.399	.360	1

Notes. Ag = Agoraphobia; An = Anxiety; De = Depression; So = Somatization; Co = Cognitive-Performance Deficits; In = Interpersonal Sensitivity; Ho = Hostility; Sl = Sleep Difficulties. Correlations $\geq .50$ are indicated in bold face.

this would result in opposite signs in the general factor, with detrimental effects on the reliability of the Anxiety subscale and the Total scale.

Discussion

The aim of this study was to examine whether the SCL-90-R can reliably measure general psychological distress as well as its multiple specific primary aspects. For the latter, we were curious to know how large the contribution was of the general factor on the subscale scores. Our results indicate that, among psychiatric outpatients, the variation in SCL-90-R total scale scores is mainly due to a strong general factor. For the most part this general factor consists of depression and anxiety symptoms. This is in concordance with the results found in previous studies (e.g., Hafkenscheid et al., 2007). Furthermore, our results indicate that Sleep Difficulties, Agoraphobia, Hostility, and Somatization, and to a lesser extent Interpersonal Sensitivity, Anxiety, and Cognitive-Performance Deficits explain some additional variance above the general factor. Depression, however, does not seem to measure anything substantial beyond general distress in this population. This means that symptoms of depression cannot be well separated from symptoms of general distress among psychiatric outpatients. The latter implies that when one uses the total scale score of the SCL-90-R, the use of the Depression subscale score seems limited.

We note that the subscale scores of the SCL-90-R are not directly related to (DSM-5) syndromes. This means that subscales that measure to a substantial extent a specific aspect of psychological distress may be of importance for different (DSM-5) syndromes. For instance, the subscale Sleep Difficulties measures a distinctive primary symptom (i.e., sleep problems) that may be of importance for the classification of multiple (DSM-5) syndromes (e.g., depression and anxiety).

Our results are in line with previous findings on the SCL-90-R in that the SCL-90-R seems to measure mainly the general factor of psychological distress (e.g., Cyr et al., 1985; Vassend & Skrondal, 1999). In addition, our results support previous findings that various subscale scores can be distinguished to a certain degree (e.g., Arrindell et al., 2006; Holcomb et al., 1983). However, this study deviates from previous studies in that the focus was not on the dimensionality of the SCL-90-R but on the unique contribution of the subscale scores of the SCL-90-R above the total scale score, and with that giving more attention to the clinical use of the subscale scores of the SCL-90-R.

We examined a Dutch sample of psychiatric outpatients. This implies that we evaluated the Dutch version of the SCL-90-R, among a rather heterogeneous sample in terms of severity and type of symptoms, with symptoms of anxiety and depression being most prevalent. The latter implies that a substantial amount of spread in scale scores can be expected, and less substantial spread on the depression and anxiety scales. The high prevalence of anxiety and mood disorders is common in clinical populations, but in a community population or specific clinical populations a higher differentiation in depression or anxiety scores may be expected. If the SCL-90-R would be assessed in a community population (e.g., for screening purposes), the depression subscale score may have more additional value above the total scale score than we found in a population of psychiatric outpatients. Future research is needed to indicate whether the results of the present study can be generalized to other (community or specific clinical) populations, and to the SCL-90-R in other languages.

Limitations and Future Research

This study was explicitly designed to evaluate the unique contribution of the subscale scores of the SCL-90-R, using the item structure of the SCL-90-R as it was designed and is currently used in clinical practice. A limitation of this approach is that possible other structures might fit the data better. For example, our results suggest that a plausible alternative structure is one where the Anxiety subscale would be split into two subscales, representing a physical and a mental component. However, a disadvantage of using such data-driven models is that they reflect to a much lesser extent the test use of the SCL-90-R in clinical practice, and therefore give results that are of less relevance for current clinical practice. Furthermore, as we verified in an additional analysis, we found comparable results when we adapted the model by allowing cross loadings between items and factors of which the modification indices were high and the item content was interpretable given the associated subscale. That is, if we added 19 cross loadings to the model, then this resulted in an improved model fit (RMSEA = .043; CFI = .904), and the results show the same sequence and division in scales with a high, moderate and low proportion of variance that is unique from the general factor.

In interpreting the additional value of a subscale above the total scale, we used particular cut-off values for the proportion of variance not attributed to the general factor. We considered the presented cut-off values useful for this rough classification. However, we acknowledge that these cut-off values are not absolute, and that one could argue about what one should consider as a substantial, moderate, or low proportion of unique variance.

Conclusion

In sum, our results suggest that clinicians can interpret and use the total scale of the SCL-90-R as an estimate of general psychological distress in a population of psychiatric outpatients. In addition, clinicians can interpret the SCL-90-R subscale scores Sleep Difficulties, Agoraphobia, Hostility, Somatization, and to a lesser extent the subscale scores Interpersonal Sensitivity, Anxiety, and Cognitive-Performance Deficits as measuring a substantial amount

of variance related to specific aspects of psychological distress in this population. However, the subscale score Depression seems to measure a very common construct as is measured with the total scale of the SCL-90-R. Because its unique variance is quite small, its usefulness as a measure of specific psychological distress symptoms beyond general psychological distress appears to be limited among the population of psychiatric outpatients.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Arrindell, W. A., Barelds, D. P. H., Janssen, I. C. M., Buwalda, F. M., & van der Ende, J. (2006). Invariance of SCL-90-R dimensions of symptom distress in patients with peri partum pelvic pain (PPPP) syndrome. *British Journal of Clinical Psychology, 45*(3), 377–391. doi: 10.1348/014466505X68924

Arrindell, W. A., & Ettema, J. H. M. (1986). *SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator* [SCL-90: Manual to a multidimensional psychopathology-indicator] (1st ed.). Lisse, The Netherlands: Swets & Zeitlinger.

Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator* [SCL-90: Manual to a multidimensional psychopathology-indicator] (2nd ed.). Amsterdam, The Netherlands: Pearson.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. doi: 10.1037/0033-2909.107.2.238

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*, 796–846. doi: 10.1111/j.1467-6494.2011.00749.x

Brunner, M., & Süβ, H.-M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement, 65*, 227–240. doi: 10.1177/0013164404268669

Cyr, J. J., McKenna-Foley, J., & Peacock, E. (1985). Factor structure of the SCL-90-R: Is there one? *Journal of Personality Assessment, 49*, 571–578. doi: 10.1207/s15327752jpa4906_2

Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring & procedures manual-II for the R(evised) version*. Baltimore, MD: Clinical Psychometric Research.

Derogatis, L. R. (1994). *SCL90-R: Administration, scoring and procedures manual*. Minneapolis, MN: National Computer Systems.

Derogatis, L. R., & Cleary, P. A. (1977). Factorial invariance across gender for the primary symptom dimensions of the SCL-90. *British Journal of Social & Clinical Psychology, 16*, 347–356.

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin, 9*, 13–28.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Scientist, 19*, 1–15.

Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., ... Urbánek, T. (2012). Testing practices in the 21st century: Developments and European psychologists' opinions. *European Psychologist, 17*, 300–319. doi: 10.1027/1016-9040/a000102

Hafkenscheid, A. J., Maassen, G. H., & Veeninga, A. T. (2007). The dimensions of the Dutch SCL-90: More than one, but how many? *Netherlands Journal of Psychology, 63*, 29–35. doi: 10.1007/BF03061059

Holcomb, W. R., Adams, N. A., & Ponder, H. M. (1983). Factor structure of the Symptom Checklist-90 with acute psychiatric inpatients. *Journal of Consulting and Clinical Psychology, 51*, 535–538. doi: 10.1037/0022-006X.51.4.535

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*, 299–331.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667–696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140. doi: 10.1080/00223891.2012.725437

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*, 23–74.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180. doi: 173-180.1991-03242-001

van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation for item scores when test data are factorially complex. *British Journal of Mathematical and Statistical Psychology, 60*, 315–337. doi: 10.1348/000711006X117574

Vassend, O., & Skrondal, A. (1999). The problem of structural indeterminacy in multidimensional symptom report instruments. The case of SCL-90-R. *Behaviour Research and Therapy, 37*, 685–701. doi: 10.1016/S0005-7967(98)00182-X

Vollebergh, W. A. M., de Graaf, R., ten Have, M., Schoemaker, C. G., Van Dorsselaer, S., Spijker, J., & Beekman, A. T. F. (2003). *Psychische stoornissen in Nederland: Overzicht van de resultaten van NEMESIS* [Psychological disorders in the Netherlands: Overview of the results of NEMESIS]. Utrecht, The Netherlands: Trimbos Instituut.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123–133. doi: 10.1007/s11336-003-0974-7

Date of acceptance: June 30, 2014
Published online: December 10, 2014

Iris A. M. Smits

Research Institute of Child Development and Education
University of Amsterdam
Nieuwe Achtergracht 127
1018 WS Amsterdam
The Netherlands
E-mail I.A.M.Smits@uva.nl