Journal of Psychosomatic Research 74 (2013) 469–478

Contents lists available at SciVerse ScienceDirect

Journal of Psychosomatic Research

Rasch scalability of the somatosensory amplification scale: A mixture distribution approach

Fabian Jasper [a,*], Urs M. Nater [b], Wolfgang Hiller [a], Ulrike Ehlert [c], Susanne Fischer [b], Michael Witthöft [a]

[a] Department of Clinical Psychology, Johannes Gutenberg University, Germany
[b] Clinical Biopsychology, Philipps University of Marburg, Germany
[c] Clinical Psychology and Psychotherapy, University of Zurich, Switzerland

ARTICLE INFO

Article history:
Received 4 October 2012
Received in revised form 23 January 2013
Accepted 12 February 2013

Keywords:
Somatosensory amplification scale
Mixture distribution analysis
Partial credit model
Rating scale model
Somatization

ABSTRACT

Objective: Somatosensory amplification refers to a person's tendency to experience somatic sensations as inappropriately intense and involves hypervigilance concerning bodily sensations. We applied the Somatosensory Amplification Scale (SSAS) in an Internet sample of young adults ($N = 3031$) to test whether the SSAS is Rasch scalable.
Methods: We applied mixture distribution extensions of the partial credit and rating scale models to identify possible subgroups that use the response set of the SSAS in different ways.
Results: A partial credit model, with two latent classes, showed a superior fit to all other models. Still, one of the SSAS items had to be removed because it showed severe underfit. Overall fit of the remaining items was acceptable, although the differentiation between at least two of the five item categories was questionable in both classes. Class 1 was characterized by a higher SSAS sum score, female gender, more somatic complaints, more anxiety, more psychosocial stress, and slightly higher depressiveness. Further exploratory analyses showed that the three mid categories of the SSAS can be collapsed without a large loss of information.
Conclusions: Our results show that a shortened version of the SSAS is Rasch scalable but also reveal that there is a lot of room for further improvements of the scale. Based on our results, Item 1 should be removed from the scale and a reduction of the number of response categories is probably warranted.

Introduction

Somatosensory amplification refers to a person's tendency to experience somatic sensations as inappropriately intense and is characterized by hypervigilance concerning bodily symptoms, which are interpreted as possible signs of an illness [1]. This kind of exaggerated attention to bodily signs plays an important role in current models of somatoform disorders and functional somatic syndromes in general [2–4] and in hypochondriasis in particular [5,6]. Although there exists a considerable theoretical and empirical overlap [7,8], hypochondriasis is considered as a nosological entity distinct from other somatoform disorders [9]. Still, self-focused attention seems to play an important role in the whole group of somatoform disorders and is a central part of current models that try to explain their etiology and maintenance [2,10–13]. Somatosensory amplification does not seem to be related to real physical hypersensitivity, because participants who show a high tendency for somatosensory amplification do not perform better in a heartbeat detection task [14–17], nor are they more sensitive to tactile stimuli [18,19]. Brown et al. [19] even reported a negative correlation between somatosensory amplification and attention to the tactile modality in an experimental design. Marcus, Gurley, Marchi, and Bauer [20] reach the conclusion that cognitive misinterpretations of bodily sensations and not real physical hypersensitivity are the reason for the relationships between measures of health anxiety and somatosensory amplification.

Up to now, we have only spoken about somatosensory amplification as a psychological construct and not about its measurement. These two aspects – the construct itself and its operationalization – are very closely related in the case of somatosensory amplification. The first questionnaire to measure somatosensory amplification consisted of five items [21,22] and was refined by Barsky et al. [1], who developed a 10-item version which showed good internal consistency ($\alpha = .82$) and retest reliability ($r = .79$). This scale was called the Somatosensory Amplification Scale (SSAS) and has been extensively used in the domain of health anxiety and hypochondriasis [5]. Speckens, Spinhoven, Sloekers, Bolk, and van Hemert [23] reported correlations between $r = .20$ and $r = .27$ for the Whiteley Index [24] and between $r = .22$ and $r = .63$ for the Illness Attitude Scales [25], depending on the particular kind of sample (outpatients, general population, etc.). One of the most recent instruments in the domain of health anxiety is the Multidimensional Inventory of Hypochondriacal Traits (MIHT), which also contains a perceptual scale that is intended to assess the "tendency to focus on bodily sensations" ([26], p. 6). Of the four scales, this one showed the highest correlation

* Corresponding author at: Department of Clinical Psychology, Johannes Gutenberg University, Mainz 55122, Germany.
E-mail address: jasper@uni-mainz.de (F. Jasper).

0022-3999/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jpsychores.2013.02.006

($r = .53$) with the SSAS. Nevertheless, the questionnaire taps important aspects of somatoform disorders and other mental disorders that involve self-awareness, which makes it a valuable instrument. Nakao and Barsky ([27], p. 1) point out that many studies have tried to clarify the role of somatosensory amplification in the pathogenesis of hypochondriasis, and that this concept "can be useful as an indicator of somatization in the therapy of a broad range of disorders, from impaired self-awareness to various psychiatric disorders."

Although there remain many questions involving the construct of somatosensory amplification, the questionnaire itself that is intended to measure this construct has not been the target of much research. Consequently, no adequate measurement model has been established for the SSAS. The pioneering study by Barsky et al. [1] successfully analyzed the questionnaire by means of classical test theory (CTT; [28]), which can be seen as a precondition for the application of a measurement theory such as item response theory (IRT; [29,30]).

IRT-based methods have been applied to various domains of clinical psychology, such as research on major depression [31–33], anxiety disorders [34] and schizophrenia [35], to name just a few. One of the first, and still the most popular model in the realm of IRT, is the so-called Rasch model for dichotomous items [36]. It assumes that the probability that a person will solve a test item (or affirm a question) depends on the difficulty of the item and the trait strength (or ability) of the person. The core of the Rasch model is the so-called logistic function, which describes this relationship between item difficulty and trait strength on the one side (often depicted on the x-axis of graphs) and the probability of a particular response on the other (often depicted on the y-axis of graphs). The most popular extension of the Rasch model is the case of polytomous items and stems from Masters [37], who introduced the *partial credit model* (PCM), which includes more than one difficulty parameter (number of item categories minus 1) for each item. These parameters are often referred to as threshold parameters and are associated with the transition from one response category to the next higher one [37]. They mark those points of the latent trait continuum at which the probability of a response in two adjacent categories (e.g., *not bothered at all*, *bothered a little*) is equal ($p = .50$). A restricted version (i.e., a sub-model) of the PCM is the *rating scale model* (RSM) by Andrich [38], which assumes that the distances between those threshold parameters are the same across all items. This restriction implies that the response set is used similarly across all items, but still allows each item to exhibit a different overall difficulty.

Compared with CTT, two popular advantages of IRT models are item parameter invariance regarding the examinee group and independency of ability scores and test difficulty [39,40]. However, the most important advantage of IRT over CTT is probably that it provides strict tests for its assumptions. Some important aspects of the measurement model are even not described at all in CTT as the type of the relationship between ability and answering probability (i.e., the item characteristic curve in IRT models). Other aspects like unidimensionality are not mentioned in the original CTT framework but some statistics like, e.g., the item-test correlation do not make sense without this assumption [28,41]. The ordinary practice to create a sum score of polytomous items for CTT-based questionnaires implies that the respondents perceive the response set as ordered along the trait continuum. By means of the application of the PCM and RSM, we can examine if this assumption is met and we are able to suggest solutions (e.g., collapsing of categories) in case it is violated. All the Rasch models mentioned above have been extended to mixture distribution models [42,43]. In a mixture distribution model, one assumes that there exist one or more latent subpopulations (i.e., classes) in which the Rasch model (or its extensions) holds, but with different model parameters [42]. This means that the item and ability parameters are calculated separately for each latent class. Thus, the mixture distribution models can be seen as a combination of latent class analysis and the Rasch model [44]. One common reason which makes the assumption of more than one latent class necessary is a different use of the response set in two or more latent classes [31,45]. Members of one class could be characterized by the tendency to avoid extreme ratings at the end of the Likert-scale whereas the members of the other class do not show such a behavior. This could mean that the members of the two classes perceive and use the response set in a different way.

Mixture distribution models are closely related to what is called differential item functioning (DIF) and plays an important role in the application of IRT [46,47]. The basic idea behind the DIF approach [47] is that the items of a test may function differently across subgroups of the target population. One example could be that the relative item difficulties are not the same for both genders (e.g., Item 1 is the easiest item within the male group but only the second easiest within the female group). The traditional way to detect DIF is to divide the sample into two distinct groups which are expected to show DIF and to calculate the IRT model in both groups separately [48]. Afterwards, one examines if the item parameters and the slope of the Item Characteristic Curves are equal across both examinee groups [49,50]. Unfortunately, this approach only functions well when the relationship between DIF and the group variable is very high [51,52]. Various studies have shown that this is rarely the case in applied research with qualitative group variables like gender, age, and race [53]. Many authors see the application of mixture distribution techniques to identify DIF as the method of choice [44,46,51,52,54].

In this article, we only focus on the Rasch model and its extensions to polytomous items and multiple classes. It is important to keep in mind that for less restrictive IRT models such as the two parameter model (or the three parameter model which involves additional guessing parameters) according to Birnbaum [55] the sum score is not a sufficient indicator of the performance of a participant. This makes their application in applied contexts problematic [56,57]. Some authors like, e.g., Scheiblechner ([56], p. 181) went so far to call those models "pseudo-Rasch models" and conclude that they are not really suitable for practical test applications.

We have pointed out that the SSAS is a widespread instrument and yet, to our knowledge, no study has examined its measurement characteristics by means of an IRT approach. The common use of SSAS sum scores implies that the categories of its items are perceived as ordered along the latent continuum. Moreover, it is desirable that the postulate of local independence holds, which means that the answer to an item only depends on a person's trait strength and the item difficulty, and not on, e.g., the answer to a preceding item [48]. If the PCM or the RSM offers a good fit to the data, one can assume that the SSAS fulfills these characteristics, which in turn increases the value of the questionnaire for future applications. Comparison of the results obtained with the PCM and RSM can help to identify whether the participants really use the response set across all items in a consistent way. Through the application of mixture distribution models, it becomes possible to search for subpopulations within our sample that are characterized by a different use of the SSAS, which could also have important implications for future applications of this instrument.

The SSAS in its original form (i.e., 10 items, 5-point Likert scale) has been applied to various domains of clinical psychology over many years. Therefore, our first aim is to examine the measurement characteristics of the SSAS in its original 10-item form by means of the mixture distribution extensions of the PCM and the RSM. When necessary, we will improve the measurement characteristics of the SSAS by the exclusion of unsuitable items and the collapsing of item categories. We hypothesize that the SSAS is Rasch scalable by means of the PCM because this model is more flexible than the RSM (i.e., perceived distances between item categories may differ across items) and the SSAS was not developed to fit the more restrictive RSM. Moreover, we are going to examine the measurement characteristics of the SSAS in a rather homogenous internet sample (i.e., compared to a population representative sample). Therefore, we do not expect that more than one latent class is necessary to provide a good fit of the model to the data.