Available online at www.sciencedirect.com

ScienceDirect

Comprehensive Psychiatry 55 (2014) 1928–1936

COMPREHENSIVE PSYCHIATRY

www.elsevier.com/locate/comppsych

Frequency vs. intensity: Framing effects on patients' use of verbal rating scale anchors

Julia Krabbe[a,b,*], Thomas Forkmann[a]

[a]Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen, Pauwelsstraße 19, 52074 Aachen, Germany
[b]Department of Surgical Intensive and Intermediate Care, University Hospital of RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany

Abstract

Background: The present study examined the hypothesis that patients' ratings would be less congruent and stable if they were asked to rate imprecise terms on frequency and intensity that were embedded in a behavioral or perceptual framework. Based on data acquired from the patients' ratings the presented terms were analyzed concerning their interindividual congruency, intraindividual stability across time and distinguishability of adjacent terms. Afterward, the results were compared to the results regarding the same analysis of unframed terms from an earlier investigation [16].
Methods: In a longitudinal design, 44 patients (age M = 39.1, SD = 15.2, 68.2% female) with a depressive disorder filled out two established questionnaires (BDI or SCL-90) and questionnaires containing frequency and intensity terms framed in sentences concerning the subjective experience of sadness. Patients should rate the terms with regard to the percentage of time or intensity that is reflected by each term at two different measuring times within one week. Data analysis contained t-tests for paired samples and effect sizes *d* according to Cohen.
Results: The congruency of framed terms was influenced by an additional factor (vocabulary skills) in comparison to unframed terms. However, congruencies for both sets of terms were rather low. In contrast to unframed items, framed terms showed no intraindividual instability for frequency and intensity terms at all, but were influenced by all of the analyzed factors (age, gender, vocabulary skills, depression, and overall mental symptom burden). Patients could distinguish more adjacent framed terms than unframed terms.
Conclusions: The results give no clear suggestion if unframed or framed terms should be preferred as verbal anchors in self-report instruments. Unframed terms seem to have a slight advantage over framed terms as they are less influenced by the patient's background. However, patients are able to distinguish more adjacent terms if presented framed in a behavioral or perceptual context they are familiar with. Frequency terms showed a higher intraindividual stability of mental representations while both groups of terms exhibited low interindividual congruency. No more than four different verbal anchors could be used safely together in rating scales, as patients with a depressive disorder would not be able to reasonably differentiate more than these.

1. Background

> "How often in the last two weeks did you have concerns about your current situation? Please choose one out of the following five options":

This could be an example for a typical item from almost every common self-report questionnaire. Those kinds of questionnaires are used not only to assess the symptoms and complaints of patients with mental disorders but also to examine subjective patient reported outcomes in all areas of health care provision [1]. Questions about the frequency of symptoms appear to be used most commonly [2].

There has been extensive research about how to design questionnaires [3–10]. When constructing a questionnaire, certain aspects have to be considered.

For example, the order of item presentation can have an impact on the patients' responses [5,11]. The items can be formulated as direct questions answered by the patient or as a statement to which the patient can indicate a certain amount of agreement or disagreement. Certainly, whether a questionnaire is constructed the one way or the other influences patients' answers [3]. The rating scales connected to the items may differ with regard to number of categories, point of origin and labeling [3,4,9,12,13].

* Corresponding author.
E-mail addresses: jkrabbe@ukaachen.de (J. Krabbe), tforkmann@ukaachen.de (T. Forkmann).

http://dx.doi.org/10.1016/j.comppsych.2014.06.010
0010-440X/

Most self-rating instruments that assess mental health are constructed based on the operational description of syndromes and diagnoses given by common diagnostic classification systems such as the International Statistical Classification of Diseases and Related Health Problems (ICD-10) or Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [14,15]. In these classificatory systems, mental disorders are defined by the presence of a set of symptoms that have been experienced by the respective patient with a given intensity and for a given time (e.g., at least for two weeks). However, when constructing questionnaires based on these definitions it has to be decided which of these two characteristics – frequency or intensity – should be used as verbal anchors for the rating scale. Both DSM-5 and ICD-10 offer no clear advice for which of the two options should be chosen.

Recently, we investigated whether we can find empirical evidence for whether frequency or intensity scales should be used in clinical self-report instruments [16]. Terms referring to some amount of frequency or intensity were evaluated concerning their interindividual congruency, intraindividual stability across time and distinguishability of adjacent terms. Data provided by a sample of 44 patients suffering from a depressive disorder showed that while overall congruency between patients was larger for intensity terms in comparison to frequency terms, frequency terms showed higher intraindividual stability. The patients were able to distinguish reliably four frequency terms and three intensity terms. Thus, these results were slightly in favor of using frequency terms when formulating verbal anchors of self-report rating scales in clinical applications and the results indicated that rating scales should have no more than four different answer categories.

These results are in line with the findings of Case's study [17] from the area of research on medical education. Members of test committees who write questions for medical examinations were asked to assign the percentage of time that was reflected by commonly used terms of frequency (e.g. always) in multiple choice questions. The results showed nearly no congruence between the participants' ratings. Thus, there is no common definition among medical professionals about the phrases used, though it was hypothesized before that there was a common definition among medical professionals.

Chang et al. [8] analyzed data from patients who suffered from cancer, stroke or HIV on two 5-point symptom self-report rating scales, each one for frequency and intensity. Using Rasch analysis they found that frequency terms provided a broader coverage of the fatigue continuum. The authors hypothesized that frequency scales outmatch intensity scales psychometrically and therefore offer a slight advantage with regard to providing a fuller coverage of the fatigue continuum.

Thus, studies investigating frequency or intensity terms as verbal anchors for self-report rating scales are rare and give no clear recommendation which to choose for questionnaire developments.

In summary, the existing research indicates that it is a rather difficult task to differentiate between frequency or intensity terms for both medical experts and patients with a depressive disorder or fatigue [8,17,18]. Interindividual congruency and intraindividual stability across time of mental representations of such terms was shown to be rather low. Moreover, the results do not give a *clear* recommendation for which scale, frequency or intensity, should be used. Thus, one might question whether usage of imprecise terms as verbal anchors of self-report rating scales is justified at all.

One might assume that a potential reason for the ambiguous pattern of results in our earlier study could have been that the patients were asked to rate imprecise terms without connecting it to a perceptual point of reference. In other words, it could be hypothesized that patients' ratings would be more congruent and stable if they were asked to rate imprecise terms that were embedded in a behavioral or perceptual framework they are familiar with. One example for such framework could be the item "I *sometimes* feel sad. Please indicate the percentage of time that is reflected by this sentence."

However, the literature concerned with this matter reports rather ambiguous findings. Schaeffer [2] hypothesized that the definition indicating the percentage of time of an item embedded in a context depends upon two aspects: the definition of the item already existing in the subject's mind and the estimated frequency of its context. In addition, the context in which the subject usually uses the presented frequency terms can have an effect on the answering pattern.

The influence of the context in which the question is embedded is depending on the subject's demographic background. For example, the subject's gender can interact with his or her mental representation of the definition of some frequency terms. In addition, age, education and ethnical background have an effect on the participants' answering behavior [19,20].

The influence of the subject's attitude and experience toward the issue covered by the questions was emphasized in an earlier investigation [21] assessing the terms selected to describe the median frequency of occurrence of an activity. Those subjects who disliked an activity assigned a higher frequency adverb to its rate of occurrence than did those who liked the activity.

So, the subjects' answering behavior can be influenced by numerous factors. Experience, attitude and usage in daily routine can affect subjects' ratings, as well as their demographic background, such as age, gender or race.

Based on the review of previous studies that highlighted numerous different variables influencing participants' answering behavior to imprecise terms, the present study examined the hypothesis that patients' ratings would be less congruent and stable if they were asked to rate imprecise terms that were embedded in a behavioral or perceptual frame. Based on data from the same patient sample as reported in Krabbe and Forkmann [16], the patients' ratings