

## Assessment of patient-reported symptoms of anxiety

Matthias Rose, MD, PhD; Janine Devine, DrPhil



*Patient self-reported symptoms are of crucial importance to identify anxiety disorders, as well as to monitor their treatment in clinical practice and research. Thus, for evidence-based medicine, a precise, reliable, and valid (ie, “objective”) assessment of the patient’s reported “subjective” symptoms is warranted. There is a plethora of instruments available, which can provide psychometrically sound assessments of anxiety, but there are several limitations of current tools that need to be carefully considered for their successful use. Nevertheless, the empirical assessment of mental health status is not as accepted in medicine as is the assessment of biomarkers. One reason for this may be that different instruments assessing the same psychological construct use different scales. In this paper we present some new developments that promise to provide one common metric for the assessment of anxiety, to facilitate the general acceptance of mental health assessments in the future.*

© 2014, AICH – Servier Research Group

Dialogues Clin Neurosci. 2014;16:197-211.

### Background

Symptoms of anxiety are part of everyone’s life. They have essential signal functions helping us to maneuver through our daily challenges. However, when anxiety is experienced without adequate stimuli, it may become a seriously burdening condition. In fact, the prevalence of anxiety disorders has been constantly rising over the past decades, becoming the seventh most burdensome condition of all diseases worldwide today.<sup>1</sup>

As with other mental disorders, patients’ self-reported symptoms are of crucial importance to diagnose anxiety disorders, as well as to monitor treatment success. For evidence-based medicine, a precise, reliable, and valid (ie, “objective”) assessment of the patient’s reported “subjective” symptoms is essential. In this paper we will focus on the state-of-the-art tools for assessing patients’ self-reported symptoms, often also called patient-reported outcomes (PROs). We favor the term “symptoms” over “outcomes,” as it includes the assess-

**Keywords:** anxiety; computerized adaptive test; item response theory; measurement; patient-reported outcome; questionnaire

**Author affiliations:** Department of Psychosomatic Medicine and Psychotherapy, Medical Clinic, Charité, Universitätsmedizin Berlin, Germany (Matthias Rose, Janine Devine); Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, USA (Matthias Rose)

**Address for correspondence:** Matthias Rose, MD, PhD, Medical Clinic for Psychosomatic Medicine, Charité, Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany (e-mail: rose@charite.de)

# Clinical research

ment of psychological constructs as outcomes, as outcome predictors, or to screen for anxiety disorders.

## Measurement of self-reported symptoms of anxiety

The empirical assessment of anxiety imposes a number of conceptual and methodological issues which need to be addressed before the use of a particular instrument can be considered.

### Conceptual issues

#### *Types of anxiety disorders*

Currently, at least four main subtypes of anxiety are usually distinguished within scientific publications, including general anxiety disorder (GAD), phobic disorders, panic disorders, and post-traumatic stress disorder (PTSD). All present with a different symptomatology, and different ethological conditions are assumed. From a measurement perspective, the first step of selecting an appropriate tool is to decide whether the assessment shall focus on more generic symptoms of anxiety, which are existent in several anxiety disorders, or more specific symptoms of one particular disorder.

#### *Dimensionality of the construct*

Instruments assessing emotional distress often sample items from different domains (eg, mood, cognition, behavior, and somatic symptoms) to capture a comprehensive set of manifest indicators of the underlying latent construct. Empirical studies have demonstrated that emotional distress can be described with a “tripartite” model, distinguishing three principal components: general distress, physiological hyperarousal, and anhedonia.<sup>2</sup> General distress is usually present in both depressive and anxiety disorders, while symptoms of anhedonia are more characteristic of depressive disorders, and symptoms of hyperarousal are more specific to anxiety disorders, in particular panic disorders and PTSD.

Accordingly, instruments developed for assessing panic disorders more often include somatic symptoms (eg, palpitations, sweating, and dyspnea) than tools primarily used for GAD or the global assessment of anxiety, which usually focus on the assessment of moods,

cognitions, or behaviors (eg, tension, nervousness, concerns, and inability to relax).

#### *Screening for anxiety disorders*

Screening for mental disorders has been frequently recommended to identify comorbid mental disorders in chronic medical conditions, such as coronary heart disease<sup>3-5</sup> or diabetes mellitus.<sup>6-9</sup> In fact, in primary care practice only about half of the patients having a depressive disorder are identified.<sup>10-12</sup> Self-reported screening tools can help the health care provider in busy daily routines to identify those patients, with minimal additional effort.<sup>13</sup>

However, the fundamental challenge of all established screening tools is that the measurement of a “dimensional” construct must support a “categorical” diagnostic decision. Thus for a natural phenomenon, like depressed mood or anxiety, a cutoff value needs to be defined above which a certain condition is likely to be classified as a pathology, according to consensus documents like the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* or *International Classification of Diseases (ICD)*.

Well-validated screening tools for depressive disorders (eg, Patient Health Questionnaire–9 [PHQ-9]) usually provide good sensitivity (ie, the likelihood that patients with depression are identified;  $\geq 0.85$ ) and specificity (ie, the likelihood that the ones being identified are in fact suffering from the disorder;  $\geq 0.85$ ).<sup>14</sup> Screening tools for anxiety disorders typically provide less favorable results, at least in clinical populations. One reason is that different types of anxiety disorders have more heterogeneous symptoms than different types of depressive disorders. Another reason is that normal anxiety reactions of clinical samples typically show a greater overlap with anxiety symptoms expressed by patients being diagnosed as having an anxiety disorder (*Figure 1*).

#### *Monitoring of anxiety symptoms*

For the treatment of mental disorders an empirical assessment of the key symptoms is essential to monitor treatment success. Symptoms are usually measured as manifest “observable” variables (eg, “In the past seven days I worried about what could happen to me” (Patient-Reported Outcomes Measurement Information System [PROMIS] Anxiety Item)<sup>15</sup> of an underlying “latent” construct (eg, anxiety). Within the majority

of established instruments, anxiety is assumed to be a state variable which can rapidly change over time. If longer recall periods are used (eg, "Indicate how much you have been bothered by... fear of dying... during the past month" (Beck Anxiety Inventory [BAI])<sup>16</sup> the manifest variable is assumed to measure a more stable aspect of the latent anxiety construct. Different recall periods may be appropriate depending on whatever treatment goal is defined.

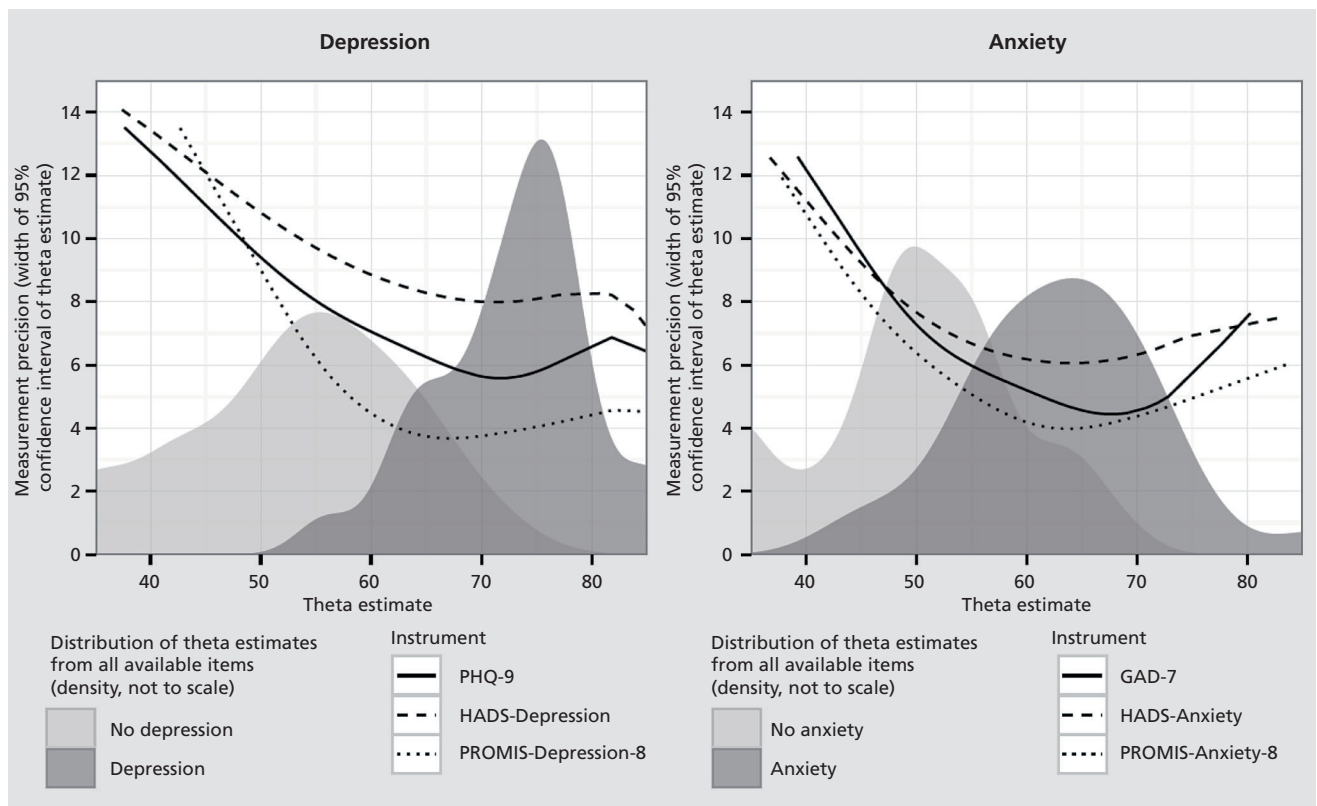
*Polarity of the construct*

From a measurement perspective we must assume a dimensional construct with lower or higher quantities. For most mental health constructs it has been extensively discussed whether the assumption of the construct being "unipolar" is appropriate (ie, from no anxiety to high

anxiety). This model is usually favored by clinicians using a pathology model where anxiety is a symptom of a disease and no symptom of the disease would be healthiness. Another model, more often favored by epidemiologists, is to assume "bipolar" mental health states, ie, a continuum from high stress resilience through situation avoidance to extreme states of anxiety. The latter conceptualizes anxiety as a natural emotional phenomenon with different states of anxiety as responses to environmental challenges. This model has a larger measurement range and allows the assessment of different anxiety levels in the normal population as well.

**Measurement issues**

Today a wide range of well validated outcomes tools is available that can be used for monitoring anxiety disorder



**Figure 1.** Example of the relation between measurement precision, measurement range of five psychometric instruments, and the distribution of the latent traits depression and anxiety in a sample of 194 heart failure patients with and without a comorbid mental disorder (for details see ref 106). PHQ, Patient Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; PROMIS, Patient-Reported Outcomes Measurement Information System

Adapted from ref 106: Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res.* 2014;23:1609-1618. Copyright © Springer Science + Business Media 2014

# Clinical research

ders. Most of these have been developed using Classical Test Theory (CTT) methods. Because self-assessment instruments have become increasingly important in the medical field, their limitations in measurement coverage and precision are more intensely discussed and other test development methods are gaining more interest. Within the next paragraphs, we will highlight some of the main restrictions of established instruments and some potential solutions to those limitations.

## *Precision and respondent burden*

The most precise and comprehensive health assessment questionnaires are rather lengthy and complex, leading to a level of respondent burden that hampers their use in routine care and often leads to substantial problems of missing data. Therefore, tools which are popular today are relatively short questionnaires.<sup>17-19</sup> They represent a compromise in measurement precision, range, and other desirable attributes in favor of practicality.<sup>20</sup> These short forms are useful for measuring the health status of large samples, but in small samples or when test scores of individual patients are evaluated, their reduced precision causes concerns.<sup>21</sup>

## *Comparability*

Another major limitation of traditional tools has been that results from different questionnaires are difficult to compare, even when two similar instruments are used to assess the same outcome of the same disease, as every instrument uses its own metric. The heterogeneity of scale-specific metrics seriously impairs comparability across study results and complicates communication among researchers and clinicians. Pooling study results from different measures in quantitative reviews or meta-analyses is difficult and may even lead to biased results.<sup>22,23</sup> The situation is as if body temperatures assessed in different settings were not comparable with one another, but were dependent on the particular thermometer used.<sup>24,25</sup>

## **New generation of measurement tools**

To make the measurement of psychological constructs more similar to biomedical ones, a standardized, efficient approach for a variety of applications, including ambulatory monitoring, clinical trials research, and population

monitoring, should be established, so that results can be compared across conditions, therapies, trials, and patients.

## *Item response theory (IRT)*

IRT provides a solution to many of the limitations of CTT. IRT methods were developed more than four decades ago,<sup>26,27</sup> and numerous attempts have been made to exploit their potential.<sup>28,29</sup> Today, IRT-based tests are well established in the educational field,<sup>30,31</sup> but have only recently been adopted in health care.<sup>32-35</sup>

Like factor analysis, IRT models assume that the measured construct is a latent variable, referred to as the IRT score ( $\theta$ ), which cannot be observed directly, but can be estimated based on responses to different items measuring the same construct. An IRT item bank consists of items measuring the same construct and a mathematical description of the items' measurement properties.<sup>36</sup> The IRT model<sup>37,38</sup> describes the probability of choosing each response to the item as a function of  $\theta$ .<sup>39,40</sup> Several different IRT models are used in health care applications, which have unique psychometric properties.<sup>30,37,40,41</sup> One important distinction of all IRT methods from CTT methods is that theta can be estimated from the responses to any subset of items in the bank.<sup>36</sup> Accordingly, researchers or clinicians can select items that are most relevant for a given group or an individual patient, and score the responses on one common metric that is independent of the choice of items. If the item bank contains items from established questionnaires, scores on these questionnaires can be predicted from estimates of  $\theta$ , even if the questionnaires themselves have not been administered ("equating").<sup>42</sup> Thus, test scores of different questionnaires can easily be compared on one common metric.<sup>43</sup>

## *Computerized adaptive tests (CATs)*

To use self-reported assessments efficiently a comprehensive electronic data capturing system is warrant. Several providers offer their services for clinical research purposes, and the market for electronic health systems, which include patient self-assessments within the electronic medical records (EMR) is currently evolving. Because modern patient self-assessments usually use some kind of computer-assisted data collection method, the application of CATs is tempting. This new generation of PRO tools promises to provide very short and reliable

assessments.<sup>30,32,41,44</sup> The principle of a CAT is to select and administer only the most informative items for every individual patient, according to her or his estimated  $\theta$  value from an IRT item bank. After each item is administered, an IRT score is reestimated to choose and apply the next best suited item for the current score estimate. CATs generally use two different ways to end the assessment (“stopping rules”): the CAT either stops after a predefined measurement precision (confidence interval) has been achieved or after a predefined total number of items have been administered. These stopping rules can also be combined or can be flexible with respect to the measurement range, eg, in a particular measurement range, a higher precision can be demanded than in other ranges. By omitting irrelevant, uninformative items, higher measurement precision is achieved while at the same time, respondent burden can be reduced.<sup>33,45-47</sup>

In 2004, the US National Institutes of Health (NIH) initiated a large project as part of their roadmap initiatives to address the need to develop a comprehensive Patient Reported Outcomes Measurement Information System (PROMIS, [www.nihpromis.org](http://www.nihpromis.org)). The initiative was launched nationwide to systematically develop a new generation of PRO tools applying IRT methods and CATs. As of today, the PROMIS initiative is the most well financed effort to improve the assessment of PROs, and is currently adopted in other languages. It will very likely influence mental health monitoring in the future, including the assessment of anxiety substantially.<sup>15,42,48,49</sup>

#### *Daily reconstruction methods and ecological momentary assessments*

Whereas the use of IRT and CAT methods have been intensely facilitated by the US NIH, the Food and Drug Administration (FDA)<sup>50</sup> has shown increasing interest in the use of Ecological Momentary Assessments (EMA) or Day Reconstruction Methods (DRM).<sup>51-53</sup> These methods address the problem of potential recall bias. It has been questioned as to what a tool measures when a patient is asked to report on his/her mental state over a time span of one or more weeks, ie, to what extent does the current mental health state distort the evaluation of the recall assessment. This may be particularly important for the assessment of emotional distress constructs like anxiety.<sup>54-56</sup> DRM and EMA typically use some kind of electronic data-capturing device (eg, a smartphone) to assess a patient's health

status under daily life conditions. The patient is asked at different planned or randomly selected intervals to report their current health status. All collected time points are later integrated for a comprehensive picture of the health status over a given time span. Although the use of EMA or DRM is infrequent until now, this may become a relevant upcoming technology for PRO assessments outside of clinical environments.

### **Self-assessment instruments of anxiety symptoms**

Searching PubMed and PubPsych using the keywords “anxiety questionnaire/survey/test/scale” in titles, revealed more than 1000 publications since the 1950s. The first anxiety questionnaires listed were developed by Cattell and Scheier<sup>57</sup> and Taylor,<sup>58</sup> both personality psychologists, who conceptualized anxiety as a trait. These anxiety measures were popular for many years after their development, but are now used infrequently.

Since the 1950s numerous anxiety questionnaires were built, first using factor analytic methods, then extending to the full range of methods used within the CTT framework, up to today's more often used IRT methods. A recent review of existing instruments shows that 145 different scales measuring anxiety are used today.<sup>15</sup>

#### **Commonly used tools**

Contemporary anxiety instruments can be categorized into two groups:

- a) generic tools, which aim to measure the common aspects of different anxiety disorders
- b) specific tools, which aim to assess anxiety in response to particular situations in the medical field (eg, dental anxiety<sup>59</sup> or cancer anxiety<sup>60-62</sup>), as well as outside the medical field (eg, test anxiety,<sup>63</sup> computer anxiety,<sup>64</sup> or dating anxiety<sup>65</sup>)

For both areas well-validated tools are available, which can be used for screening purposes, or outcome assessments, or both (*Tables I and II*).

#### *Generic anxiety measures*

One prominent example of a generic instrument to monitor anxiety symptoms is the State Trait Anxiety Inventory (STAI).<sup>66</sup> Its items were primarily developed to optimize the psychometric properties of the tool.

# Clinical research

Instrument	Domains	Number of items	Recall period	Time to complete	Psychometric properties
<b>&gt;1000 citations</b>					
State Trait Anxiety Inventory (STAI) <sup>66,107</sup>	Anxiety (state and trait)	20 state and 20 trait items	Currently and generally	4-8 min per scale	Good psychometric properties (internal consistency, 0.86-0.95; retest reliability, 0.65-0.89; proven validity: sensitivity, 0.82; specificity, 0.88), short versions. Norm data available.
Hospital Anxiety and Depression Scale (HADS) <sup>17,19</sup>	Anxiety and depression	7 anxiety and 7 depression items	Past week	2-5 min	Good psychometric properties (internal consistency, 0.76-0.80; retest reliability 0.70; sensitivity and specificity for anxiety disorders, 0.85). Norm data available.
<b>100-1000 citations</b>					
Generalized Anxiety Disorder-7 (GAD-7) <sup>18,108</sup>	Anxiety (items reflect DSM-IV criteria for GAD)	7 anxiety items	Over the last 2 weeks	5 min	Good psychometric properties (internal consistency, 0.89; good reliability and convergent validity: sensitivity, 0.80; specificity, 0.86). Norm data available.
Beck Anxiety Inventory (BAI) <sup>16</sup>	Anxiety (cognitive and somatic components)	21 anxiety items	Last week	5-10 min	Good psychometric properties (internal consistency, >0.9; retest reliability, 0.6-0.9; correlation to HAMRS, 0.5; STAI, 0.5; BDI-II, r=0.66; SCL-90, 0.8; responsive: sensitivity, 0.67, specificity, 0.93). Norm data available.
Zung Anxiety Scale <sup>109</sup>	Anxiety (cognitive, autonomic, motor, central nervous system symptoms)	20 anxiety items	During the past several days	10-15 min	Moderate psychometric properties (internal consistency, 0.74-0.77) discriminates well between patients diagnosed with and without anxiety disorders; correlation between Zung and BDI, 0.59.
<b>&lt;100 citations</b>					
Mood and Anxiety Symptoms Question. (MASQ) <sup>2,110,111</sup>	Tripartite model (general distress, anhedonia, hyperarousal)	90 items (short form: 30 items)	Past week	10 min short version:	Good psychometric properties (good internal consistency, >0.87; good validity). Norm data available.
Patient Health Questionnaire-4 (PHQ-4) <sup>20,67</sup>	Anxiety and depression	2 anxiety and 2 depression items	Last 2 weeks	2 min	Moderate psychometric properties for anxiety scale (internal consistency, 0.75; sensitivity, 0.86, specificity, 0.70). Norm data available.
Penn State Worry Question <sup>88,112</sup>	Worries	16 worry items	Current	10-15 min	Good psychometric properties (high internal consistency; good test-retest reliability; good discriminant validity GAD).
Anxiety Screening Question. (ASQ-15) <sup>113</sup>	Anxiety disorders (panic disorders, and GAD)	15 anxiety items	Last week	10 min	Good psychometric properties (retest reliability, 0.6; sensitivity, >0.82; specificity, >0.70 for GAD).
Anxiety Disorder Diagnostic Question. (ADDQ) <sup>114</sup>	Fear, anxiety/worry, escape/avoidance behaviors, physiological, and distress symptoms, interference	8 anxiety questions, 1 symptom list, 3 open questions	Current	10 min	Good psychometric quality (good internal consistency; convergent and discriminant validity; sensitive to change).

**Table I.** Examples of generic anxiety questionnaires. Further examples of generic anxiety scales are: Cattell and Scheier's Anxiety Scale,<sup>57,115</sup> Taylor Manifest Anxiety Scale,<sup>58</sup> Worry and Anxiety Questionnaire,<sup>116</sup> Lehrer Woolfolk Anxiety Symptom Questionnaire (LWASQ),<sup>117</sup> Four Systems Anxiety Questionnaire,<sup>118</sup> Worries and emotionality Scale,<sup>119</sup> Most anxiety questionnaires have been built based on principles of the classical test theory (CTT), however some anxiety tests have also been reanalyzed using modern item response theory (IRT) methods.<sup>107,120,121</sup>

Strengths	Weaknesses	Screening	Royalty free	Obtainable
Among the most widely researched and used measures, offered in 48 languages. State scale is sensitive to the detection of longitudinal change.	Relatively long instrument, high correlations between state- and trait-scale.	x		Copyright: Mind Garden, 855 Oak Grove Avenue, Suite 215, Menlo Park, CA 94025, www.mindgarden.com/index.htm
Very widely used screening measure, offered in various languages, short screener to detect the presence of clinically significant symptoms covering tension, worry, fear, panic, difficulty in relaxing, and restlessness.	Some evidence of reduced validity in the elderly.	x		Copyright: Nfer Nelson, The Chiswick Centre, 414 Chiswick High Road, London, W4 5TF, UK, www.nfer-nelson.co.uk
Offered in different languages, cut-off scores for GAD available.	Screening for GAD diagnosis only.	x	x	www.phqscreeners.com
Developed to minimize the overlap between depression and anxiety scales, youth-specific BAI available.	Focus on somatic aspects (eg, heart racing, dizziness) may overrate anxiety in medical conditions.	x		Copyright: Pearson Assessment www.pearsonassessments.com
Frequently replicated psychometric results.	Several Zung items have higher correlations with the BDI than with the total Zung score.		x	www.psychology-tools.com/zung-anxiety-scale/
Strong theoretical model.	Long version does not fit the 3 factor model very well.		x	Contact author
Ultra-short screener for depression and anxiety.	PHQ-2 and PHQ-9 for depression measurement PHQ-4 is not well accepted yet, however the 2 depression items, part of the PHQ-4, are widely used (PHQ-2).	x	x	Available for free in multiple languages from www.phqscreeners.com
Detailed assessment of worries.	Restricted to worries.		x	www.outcometracker.org/library/PSWQ.pdf
Tested against a standardized clinical interview (CIDI).	Specificity is only sufficient for DSM-IV GAD.	x		Contact author
Brief four-section index.	Not reported.	x	x	www.midss.org/sites/default/files/addq.pdf

(Continued) Anxiety scales are often combined with measuring depression, for a more extensive overview of 34 tests, measuring anxiety and depression combined see reference 122; an example of a frequently used clinician rating scale is the Hamilton Anxiety Scale (HAM-A; <http://www.psychiatrytimes.com/clinical-scales-anxiety/ham-hamilton-anxiety-scale>).<sup>123</sup> GAD, General Anxiety Disorder; HADS, Hospital Anxiety and Depression Scale; PHQ, Patient Health Questionnaire; PROMIS, Patient-Reported Outcomes Measurement Information System

# Clinical research

Twenty items assess “how you feel right now, at this moment,” and twenty items measure “how you generally feel,” with the intention to differentiate between anxiety states and traits. Both scales have a good internal constancy with a Cronbach  $\alpha > 0.90$ . However, as the Cronbach  $\alpha$  value is dependent on the number of items of the scale, long scales usually have high Cronbach  $\alpha$  values. Both STAI scales show a high correlation to depression scales, as the majority of items capture a “negative mood” or “general distress” aspect (eg, feeling tense, upset, frightened, indecisive, strained, etc) which is also assessed by typical depression scales.

An example of a shorter more recently published tool with a clinical background is the GAD-7.<sup>18</sup> The item set of this instrument was primarily developed to capture the anxiety construct as defined by the *DSM-IV*. Although the GAD-7 only uses one third the number of items as the STAI, its psychometric properties are almost as good.

Some tests, like the Hospital Anxiety and Depression Scale (HADS), provide one scale for anxiety and one for depression in the same instrument to address the issue of content overlap. The HADS provides good psychometric criteria for both scales,<sup>19</sup> but how distinct both scales are is influenced by the populations being studied. The PHQ-4 is an example of an ultra-short screening tool, which also allows for assessing anxiety and depression by one tool.<sup>67</sup> However, the PHQ-4 with two items per scale is less precise, useful in large epidemiological studies, but not well suited for smaller studies or individual clinical decision making.

The BAI is a scale, which aims to measure aspects of anxiety that are most distinct from the depression construct. Thus, the BAI focuses more on the evaluation of “hyperarousal” (eg, heart pounding/racing, hands trembling) targeting more the physical symptoms of anxiety. Thus, the BAI is prone to be used to assess panic, phobic, or PTSD disorders.

A few instruments, like the Anxiety Screening Questionnaire–15 (ASQ-15) or GAD-7, were primarily constructed to screen for anxiety disorders. As we previously stated, self-assessment tools used to identify anxiety disorders are often show less favorable psychometric characteristics than screening tools for depressive disorders. Nevertheless, if screening tools are used carefully, they can still provide valuable information for the primary care provider. Tools like the GAD-7,<sup>18</sup> which are primarily developed for screening purposes, are also valuable, responsive “outcome” tools. Furthermore, for many “outcome” measures, like the HADS or STAI, cut-off scores have been established to also allow screening for anxiety disorders, with good sensitivity and specificity (*Table I*).

## Specific anxiety measures

Most instruments assessing specific anxiety disorders evaluate some kind of social phobia, like the Social Phobia Inventory (SPIN) or the Social Anxiety Questionnaire (SAQ-A30). However, for a huge number of other specific phobias validated instruments exist, such

Instrument	Domains	Number of items	Recall period	Time to complete	Psychometric properties
<b>&lt;100 citations</b>					
Social Interaction Scale (SIAS) and Social Phobia Scale (SPS) <sup>88,124-126</sup>	Two companion measures for social phobia fears: fear of being scrutinized and fear of general social interaction.	20 items per scale	Current	10-25 min	High internal consistency and retest reliability, good convergent and discriminant validity, sensitive to change.
Social Phobia Inventory (SPIN) <sup>127,128</sup>	Main spectrum of social phobia such as fear, avoidance, and physiological symptoms.	17 items	During the past week	10-15 min	Good internal consistency, retest reliability, convergent and divergent validity, sensitive to change, cut-off scores available.
Social Anxiety Questionnaire (SAQ-A30) <sup>129</sup>	Social phobia/anxiety structured in five dimensions.	30 items	Current	15-20 min	Well-proven factor structure, good internal consistency, construct validity, cultural invariance, cut-off scores available.

**Table II.** Examples of specific anxiety questionnaires. An example of a frequently used clinician rating scale is the Liebowitz Social Anxiety Scale (LSAS; <http://healthnet.umassmed.edu/mhealth/LiebowitzSocialAnxietyScale.pdf>).<sup>130,131</sup> BDI, Beck Depression Inventory; DSM-IV, Diagnostic Statistical Manual of Mental Disorders IV; HAMRS, Hamilton Depression Rating Scale



as the Dental Anxiety Questionnaire,<sup>59</sup> Anxiety about Death Questionnaire,<sup>68,69</sup> Cardiac Anxiety Questionnaire,<sup>70</sup> Burn Specific Anxiety Scale,<sup>71,72</sup> Preoperative Anxiety Questionnaire,<sup>73,74</sup> Anxiety and Preoccupation of Sleep Questionnaire,<sup>75</sup> Prostate Cancer-specific Anxiety Scale,<sup>76</sup> Radiotherapy Categorical Anxiety Scale,<sup>61</sup> Dyspnea-Related Anxiety,<sup>77,78</sup> COPD Anxiety Questionnaire,<sup>79</sup> Pain Anxiety Questionnaire(s),<sup>80</sup> Glasgow Anxiety Scale for People with an Intellectual Disability,<sup>81</sup> Florida Shock Anxiety Scale,<sup>82</sup> Faces Anxiety Scale for Intensive Care Patients,<sup>83</sup> Optometric Patient Anxiety Scale,<sup>84</sup> Pregnancy Anxiety Scale,<sup>85</sup> or Psychotic Anxiety Scale.<sup>86</sup>

Examples of specific anxiety scales used outside of the medical field include the Flight Anxiety Situations Questionnaire,<sup>87</sup> Separation Anxiety Symptoms Inventory,<sup>88</sup> Anxiety Control Questionnaire,<sup>89</sup> A/D Goldberg Questionnaire: Anxiety and Depression at Work,<sup>90</sup> Test Anxiety Questionnaire,<sup>63</sup> Mathematics Anxiety Scale,<sup>91</sup> Statistical Anxiety Scale,<sup>92</sup> Job Anxiety Scale,<sup>93</sup> Social Physique Anxiety Scale,<sup>94</sup> Equine Anxiety Questionnaire,<sup>95</sup> Dating Anxiety Scale,<sup>65</sup> Anxiety Scale for Music Students,<sup>96</sup> and Computer Anxiety Scale,<sup>64</sup> among many others.

**New measures**

In addition to the numerous established tools, a few new instruments developed using IRT methods have emerged, which have the potential to significantly im-

prove self-reported anxiety measures and to provide a common metric for many existing tools.<sup>33,42</sup>

The first IRT item bank for anxiety was published a decade ago in Germany. It showed that fourteen different tools provide shared information which can be scored on one common metric. This IRT item bank was used to build the first CAT for the assessment of anxiety (A-CAT),<sup>97-99</sup> which has improved psychometric characteristics compared with established tools of similar length, and has been implemented in clinical practice ever since.<sup>100</sup>

Other groups have recently published a CAT for Anxiety in the US (CAT-ANX).<sup>101</sup> Probably the most advanced IRT item bank today is provided by the PROMIS initiative.<sup>15</sup> After intensive qualitative work examining established static anxiety tests of different lengths a new large anxiety item bank was built during an extensive quantitative development process. The PROMIS anxiety item bank and CAT are publicly available, royalty free, at the Assessment Center ([www.assessmentcenter.net/](http://www.assessmentcenter.net/)). Simulation studies show very favorable results<sup>15</sup> and results from real CAT applications are expected soon. For an overview of available CATs for anxiety see *Table III*.

**Choosing the “right” instrument**

An essential question is how to choose the best anxiety instrument out of the large set of available tools. Obviously there is no best instrument for all research ques-

Strengths	Weaknesses	Screening	Royal-ty free	Obtainable
Easy to score.	Restricted to social phobia diagnosis only.	x	x	<a href="http://www.academia.edu/">www.academia.edu/</a>
Mini-SPIN with only 3 questions available.	Restricted to social phobia diagnosis only.	x	x	<a href="http://psychology-tools.com/spin/">http://psychology-tools.com/spin/</a>
Test developed based on several years of work by the research team in 18 Latin-American countries, Spain, and Portugal. Cross-culturally tested.	Restricted to social phobia diagnosis only.	x	x	<a href="http://www.midss.org/sites/default/files/saq-a30_english.pdf">www.midss.org/sites/default/files/saq-a30_english.pdf</a>

**Table II.** Continued

# Clinical research

Instrument	Domains		Number of items	Recall period	Time to complete
<b>&lt;100 citations</b>					
A-CAT <sup>97-99</sup>	Anxiety (unidimensional bipolar).	2 348 clinical patients with/without different anxiety disorders	50 items (full bank) 6 items (CAT)	current up to 4 weeks	1.7 ± 1.1 min with precision based stopping rule
CAT-ANX <sup>101</sup>	Anxiety (multi-dimensional: mood, cognition, behavior, somatization)	1 614 clinical patients with/without GAD	431 items (full bank) 12 items (CAT)	2 weeks	2.5 ± 1.6 min
CAT of the Mood and Anxiety Spectrum Scales <sup>132</sup>	Anxiety and depression (bi-factor model general factor: anxiety, somatic complaints).	800 clinical patients	626 items (full bank), 24 items (CAT)	Not reported	Not reported
PROMIS Anxiety-CAT <sup>15</sup>	Anxiety (unidimensional unipolar).	>15 000 mainly general population internet sample	29 items (full bank), number of CAT items vary	7 days	Not reported

**Table III.** Available computerized adaptive tests (CATs) for anxiety. CES-D HADS, Hospital Anxiety and Depression Scale; MASQ, Mood and Anxiety Symptom Questionnaire; STAI, State-Trait Anxiety Inventory

tions, however, the following considerations may guide one's decision:

- In our opinion the main concern should be whether the content of the items captures the content of interest. One should be cautious about the title of the instrument or the scales, as those sometimes misrepresent the items.
- A second thought should be given to the relation between measurement precision and respondent burden. In general, small sample sizes (or the use of anxiety measures in clinical practice) require higher measurement precision than larger sample sizes, eg, for epidemiological studies. Essentially, higher measurement precision requires more items and leads to higher respondent burden if static tests are used. In most cases, a compromise between precision and respondent burden needs to be found. Individually tailored, dynamic tools (CATs) can provide shorter and more precise measures, but their practical advantages need to be demonstrated.
- Third, measurement precision should be considered in relation to measurement range and the distribution of the sample being studied. Pilot studies are useful to identify floor or ceiling effects (a high proportion of respondents at the scale extremes), which compromise the interpretation of study results. Traditional psychometric characteristics, like

Cronbach's  $\alpha$ , do not provide information in this respect, as the precision of an instrument is dependent on the measurement range. Thus, we find the term "reliable range" more useful. Modern psychometric methods make the relationship between measurement range and precision more transparent, and may provide a rational guide to prefer one tool above the other (see *Figure 1*; for more details see ref 102).

- Last but not least, any tools considered should have a manual or an article describing the development sample, the psychometric model and methods used for development, as well as information about psychometric properties of the test, including reliability and validity results. Norm samples of the test are useful for interpretation and to compare results across studies.

Many self-assessment tools can be found online, together with test descriptions (eg, [www.proqolid.org](http://www.proqolid.org) or [www.psychometrikon.de](http://www.psychometrikon.de)).

## Modes of assessment

As important as the instrument being considered, in our opinion, is the mode of assessment for a successful use of self-assessment tools. In general the requirements for the use of PRO tools differ between clinical

Psychometric properties	Strengths	Weaknesses	Screening	Royal-ty free	Obtainable
Convergent and discriminant validity shown: correlation with legacy tools: 0.60 STAI, 0.66 HADS (anxiety scale).	First CAT built to measure anxiety, integrated into clinical routines for a decade.	Inconsistent recall periods as it combines items from different existing measures for a common metric.		x	Contact authors
CAT was strongly related to GAD diagnosis: sensitivity, 0.65; specificity, 0.93; CAT-DI /CAT-ANX: 0.82.	Based on review of 100 depression and anxiety scales.	Low sensitivity for GAD long CAT.	x		Contact authors
Correlation to legacy tools not reported.	Hierarchical bifactor model.	Longest CAT.			Contact authors
correlation to MASQ, 0.80; CES-D, 0.75.	Based on review of 145 anxiety scales, extensive state-of-the-art qualitative and quantitative item bank development.	High correlation between anxiety and depression PROMIS item banks, 0.81.		x	<a href="http://www.assessmentcenter.net/">www.assessmentcenter.net/</a> <a href="http://www.nihpromis.org">www.nihpromis.org</a>

Table III. Continued

practice and research. For clinical practice settings, an individual report must be provided without delay to be useful for clinical decision making. For clinical research aggregated reports or data banks are usually sufficient for a study.

We have developed and used different electronic-PRO (ePRO) systems at our clinic for almost two decades and we believe that only electronic data collection methods meet the clinical standards in busy routine care. In addition ePRO systems are less expensive than paper-pencil assessments, as they reduce staff time for administration, scoring, and report preparation substantially.<sup>103</sup>

For clinical research purposes this decision is more complex. First of all, it depends if an electronic data capturing system is already available, or an open source (eg, [www.limesurvey.org](http://www.limesurvey.org)) or a commercial system (eg, [www.unipark.de](http://www.unipark.de)) can be used, which meets data protection requirements. Second, electronic assessments are often favored as they reduce missing data and avoid false data entries, but if sample sizes are small, paper-pencil assessments may be more convenient and less expensive than programming an assessment. However, other differences between electronic and paper-pencil assessment may be considered as well, and have been discussed elsewhere in detail.<sup>104</sup> From a psychometric

perspective many modes of assessment can be used interchangeably, including paper-pencil assessment, computer assessments on desktop computers or smartphones, as well as interactive voice-recognition (IVR), whereas telephone interviews are typically more biased by social desirability.<sup>105</sup>

## Outlook

The empirical assessment of anxiety and other mental health symptoms is essential for evidence-based medicine. There is a plethora of instruments available, which can provide valid and reliable assessments of anxiety. However, the measurement of psychological constructs is not as established as the measurement of biomarkers. One reason for this may be that all instruments provide different scores, making intuitive interpretation and communication more difficult.

For a greater acceptance in the medical field, beyond mental health care, we believe we need to move away from an instrument-defined measurement and towards a construct-defined measurement system. New developments, such as IRT and CATs, make different instruments easily comparable by a standardized common metric, like different thermometers measuring on the same temperature scale. □

# Clinical research

**Acknowledgements:** We like to acknowledge the valuable, thoughtful comments and careful edits from our colleague and coworker Barbara Gandek, from the Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA, as well as the contribution from Felix Fischer, from the Medical Clinic for Psychosomatic Medicine, Charité, Universitätsmedizin Berlin, who shared his

work for *Figure 1* with us for this paper. This work was supported by the German Research Society (Deutsche Forschungsgemeinschaft DFG RO 2258/2-1, PI Rose), and a NIMH grant (R01MH082953, PI Rose). It was also supported by the University of Massachusetts, Medical School, Worcester, MA, USA, and the Charité, Universitätsmedizin Berlin, Germany.

## **Evaluación de los síntomas de ansiedad percibidos por el paciente**

*Los síntomas auto-informados por el paciente son de crucial importancia para identificar los trastornos de ansiedad, como también para monitorizar su tratamiento en la práctica clínica y en investigación. Por lo tanto, para la medicina basada en la evidencia está justificada la evaluación precisa, confiable y válida (es decir, "objetiva") de los síntomas "subjetivos" percibidos por el paciente. Existe una gran cantidad de instrumentos disponibles que pueden proporcionar evaluaciones psicométricas fidedignas de la ansiedad, pero hay algunas limitaciones de las herramientas actuales que necesitan ser consideradas con cuidado para un empleo exitoso. Sin embargo, en medicina, la evaluación empírica del estado de salud mental no es tan aceptada como ocurre con los biomarcadores. Una razón para esto puede ser que los diferentes instrumentos que evalúan el mismo constructo psicológico utilizan distintas escalas. En este artículo se presentan algunos nuevos desarrollos que prometen proporcionar una métrica común para la evaluación de la ansiedad, y así facilitar a futuro la aceptación general de las evaluaciones en salud mental.*

## **Évaluation des symptômes anxieux rapportés par les patients**

*Les symptômes rapportés par les patients sont très importants pour identifier les troubles anxieux et surveiller leur traitement en recherche et en pratique cliniques. Une évaluation précise, fiable et valable (c'est-à-dire « objective ») des symptômes « subjectifs » rapportés par le patient est donc justifiée pour la médecine basée sur les preuves. Les instruments disponibles sont très nombreux, capables d'évaluer l'anxiété de façon psychométrique mais certaines limites d'outils actuels méritent d'être soigneusement analysées pour une utilisation fructueuse. L'évaluation empirique de l'état de santé mentale n'est néanmoins pas aussi bien acceptée en médecine que l'évaluation des biomarqueurs, ce qui peut s'expliquer par l'utilisation d'instruments différents pour évaluer la même construction psychologique à l'aide d'échelles différentes. Nous présentons dans cet article de nouvelles avancées promettant la conception d'une mesure commune d'évaluation de l'anxiété afin de faciliter à l'avenir l'acceptation générale de l'évaluation de la santé mentale.*

## REFERENCES

1. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2163-2196.
2. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol*. 1991;100:316-336.
3. Albus C, Jordan J, Herrmann-Lingen C. Screening for psychosocial risk factors in patients with coronary heart disease—recommendations for clinical practice. *Eur J Cardiovasc Prev Rehabil*. 2004;11:75-79.
4. Janeway D. An integrated approach to the diagnosis and treatment of anxiety within the practice of cardiology. *Cardiol Rev*. 2009;17:36-43.
5. Schussler G, Heuft G. [Anxiety and depression in patients with medical diseases]. *Z Psychosom Med Psychother*. 2008;54:354-367.
6. Bernstein CM, Stockwell MS, Gallagher MP, Rosenthal SL, Soren K. Mental health issues in adolescents and young adults with type 1 diabetes: prevalence and impact on glycemic control. *Clin Pediatr (Phila)*. 2013;52:10-15.
7. Hall PA, Coons MJ, Vallis TM. Anxious temperament and disease progression at diagnosis: the case of type 2 diabetes. *Psychosom Med*. 2008;70:837-843.
8. McHale M, Hendrikz J, Dann F, Kenardy J. Screening for depression in patients with diabetes mellitus. *Psychosom Med*. 2008;70:869-874.
9. Pouwer F. Should we screen for emotional distress in type 2 diabetes mellitus? *Nat Rev Endocrinol*. 2009;5:665-671.
10. Nuyen J, Volkens AC, Verhaak PF, Schellevis FG, Groenewegen PP, Van den Bos GA. Accuracy of diagnosing depression in primary care: the impact of chronic somatic and psychiatric co-morbidity. *Psychol Med*. 2005;35:1185-1195.
11. Seelig MD, Katon W. Gaps in depression care: why primary care physicians should hone their depression screening, diagnosis, and management skills. *J Occup Environ Med*. 2008;50:451-458.
12. Sherbourne CD, Wells KB, Hays RD, Rogers W, Burnam MA, Judd LL. Subthreshold depression and depressive disorder: clinical characteristics of general medical and mental health specialty outpatients. *Am J Psychiatry*. 1994;151:1777-1784.
13. US Preventive Task Force (USPSTF). Screening for Depression. Recommendations and Rationale. <http://www.uspreventiveservicestaskforce.org/3rduspstf/depression/depressrr.htm>. Accessed April 18, 2014.

14. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606-613.
15. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. *Assessment.* 2011;18:263-283.
16. Beck A, Steer R. *Beck Anxiety Inventory Manual.* San Antonio, TX: Harcourt Brace and Company; 1993.
17. Snaith RP. The Hospital Anxiety And Depression Scale. *Health Qual Life Outcomes.* 2003;1:29.
18. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166:1092-1097.
19. Zigmond A, Snaith R. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67:361-370.
20. Lowe B, Wahl I, Rose M, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J Affect Disord.* 2010;122:86-95.
21. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995;4:293-307.
22. Davidson KW, Kupfer DJ, Bigger JT, et al. Assessment and treatment of depression in patients with cardiovascular disease: National Heart, Lung, and Blood Institute Working Group Report. *Psychosom Med.* 2006;68:645-650.
23. Puhan MA, Soesilo I, Guyatt GH, Schunemann HJ. Combining scores from different patient reported outcome measures in meta-analyses: when is it justified? *Health Qual Life Outcomes.* 2006;4:94.
24. Ware JE Jr. Improvements in short-form measures of health status: introduction to a series. *J Clin Epidemiol.* 2008;61:1-5.
25. Ware JE. Measuring patients' views: the optimum outcome measure. *BMJ.* 1993;306:1429-1430.
26. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores.* Reading, UK: Addison-Wesley; 1968.
27. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests.* Chicago, IL: University of Chicago Press; 1960.
28. Bech P, Allerup P, Rosenberg R. The Marke-Nyman temperament scale. Evaluation of transferability using the Rasch item analysis. *Acta Psychiatr Scand Suppl.* 1978;57:49-58.
29. Fisher AG. The assessment of IADL motor skills: an application of many-faceted Rasch analysis. *Am J Occup Ther.* 1993;47:319-329.
30. Wainer H, Dorans NJ, Eignor D, et al. *Computerized Adaptive Testing: A Primer.* 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 2000.
31. Cohen Y, Ben-Simon A, Tractinsky N. *Computerized Adaptive Test of English Proficiency.* Jerusalem, Israel: 1989.
32. Ware JE, Jr, Kosinski M, Bjorner JB, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res.* 2003;12:935-952.
33. Cella D, Chang CH. A discussion of item response theory and its applications in health status assessment. *Med Care.* 2000;38:1166-1172.
34. Ware JE, Jr, Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing: a brief summary of ongoing studies of widely used headache impact scales. *Med Care.* 2000;38:1173-1182.
35. Bjorner JB, Kosinski M, Ware JE Jr. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HITTM). *QLR.* 2003;12:913-933.
36. Bjorner JB, Kosinski M, Ware JE Jr. Computerized adaptive testing and item banking. In: Fayers PM, Hays RD, eds. *Assessing Quality of Life.* 2nd ed. Oxford, UK: Oxford University Press; 2004.
37. van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory.* Berlin, Germany: Springer; 1997.
38. Fischer GH, Molenaar IW. *Rasch Models - Foundations, Recent Developments, and Applications.* 1st ed. Berlin, Germany: Springer-Verlag; 1995.
39. Embretson SE. The New Rules of Measurement. *Psychological Assessment.* 1996;8:341-349.
40. Embretson SE, Reise SP. *Item Response Theory for Psychologists.* London, UK: Lawrence Erlbaum Associates Inc; 2000.
41. van der Linden WJ, Glas CAW. *Computerized Adaptive Testing: Theory and Practice.* Dordrecht, the Netherlands: Kluwer Academic Publishers; 2000.
42. Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: Linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *J Anxiety Disord.* 2013;28:88-96.
43. Bjorner JB, Kosinski M, Ware JE Jr. Using item response theory to calibrate the Headache Impact Test (HIT) to the metric of traditional headache scales. *Qual Life Res.* 2003;12:981-1002.
44. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res.* 1997;6:595-600.
45. Hambleton RK, Slater SC. Item Response Theory Models and Testing Practices: Current International Status and Future Directions. *Eur J Psychol Assess.* 1997;13:21-28.
46. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care.* 2000;38:1128-1142.
47. McDonald RP. Future directions for item response theory. *Int J Educ Res.* 1989;13:205-220.
48. Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res.* 2010;19:595-607.
49. Irwin DE, Gross HE, Stucky BD, et al. Development of six PROMIS pediatric proxy-report item banks. *Health Qual Life Outcomes.* 2012;10:22.
50. Food and Drug Administration. *Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Rockville, MD: FDA; 2009.
51. Dockray S, Grant N, Stone AA, Kahneman D, Wardle J, Steptoe A. A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the Day Reconstruction Method. *Soc Indic Res.* 2010;99:269-283.
52. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. *Science.* 2004;306:1776-1780.
53. Stone AA, Schwartz JE, Broderick JE, Deaton A. A snapshot of the age distribution of psychological well-being in the United States. *Proc Natl Acad Sci U S A.* 2010;107:9985-9990.
54. Edmondson D, Shaffer JA, Chaplin WF, Burg MM, Stone AA, Schwartz JE. Trait anxiety and trait anger measured by ecological momentary assessment and their correspondence with traditional trait questionnaires. *J Res Pers.* In press.
55. Kashdan TB, Collins RL. Social anxiety and the experience of positive emotion and anger in everyday life: an ecological momentary assessment approach. *Anxiety Stress Coping.* 2010;23:259-272.
56. Pfaltz MC, Michael T, Grossman P, Margraf J, Wilhelm FH. Instability of physical anxiety symptoms in daily life of patients with panic disorder and patients with posttraumatic stress disorder. *J Anxiety Disord.* 2010;24:792-798.
57. Cattell RB. The nature and measurement of anxiety. *Sci Am.* 1963;208:96-104.
58. Taylor J. The Taylor manifest anxiety scale and intelligence. *J Abnorm Psychol.* 1955;51:347.
59. Appukkuttan DP, Tadepalli A, Cholan PK, Subramanian S, Vinayagavel M. Prevalence of Dental Anxiety among Patients Attending a Dental Educational Institution in Chennai, India - A Questionnaire Based Study. *Oral Health Dent Manag.* 2013;12:289-294.
60. Dale W, Hemmerich J, Meltzer D. Extending the validity of the Memorial Anxiety Scale for Prostate Cancer (MAX-PC) at the time of prostate biopsy in a racially-mixed population. *Psychooncology.* 2007;16:493-498.
61. Shimotsu S, Karasawa K, Kawase E, et al. An investigation of anxiety about radiotherapy deploying the Radiotherapy Categorical Anxiety Scale. *Int J Clin Oncol.* 2010;15:457-461.
62. van Beek MH, Voshaar RC, van Deelen FM, Van Balkom AJ, Pop G, Speckens AE. The cardiac anxiety questionnaire: cross-validation among cardiac inpatients. *Int J Psychiatry Med.* 2012;43:349-364.
63. Sarason SB, Gordon EM. The test anxiety questionnaire: scoring norms. *J Abnorm Psychol.* 1953;48:447-448.
64. Lester D, Yang B, James S. A short computer anxiety scale. *Percept Mot Skills.* 2005;100:964-968.

# Clinical research

65. Glickman AR, La Greca AM. The Dating Anxiety Scale for Adolescents: scale development and associations with adolescent functioning. *J Clin Child Adolesc Psychol.* 2004;33:566-578.
66. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press; 1983.
67. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50:613-621.
68. Peters L, Cant R, Payne S, et al. Emergency and palliative care nurses' levels of anxiety about death and coping with death: a questionnaire survey. *Australas Emerg Nurs J.* 2013;16:152-159.
69. Templer DI. The construction and validation of a Death Anxiety Scale. *J Gen Psychol.* 1970;82:165-177.
70. Sardinha A, Nardi AE, Araujo CG, Ferreira MC, Eifert GH. Brazilian portuguese validated version of the cardiac anxiety questionnaire. *Arg Bras Cardiol.* 2013;101:554-561.
71. Taal LA, Faber AW. The burn specific pain anxiety scale: introduction of a reliable and valid measure. *Burns.* 1997;23:147-150.
72. Taal LA, Faber AW, van Loey NE, Reynders CL, Hofland HW. The abbreviated burn specific pain anxiety scale: a multicenter study. *Burns.* 1999;25:493-497.
73. Kain ZN, Mayes LC, Cicchetti DV, Bagnall AL, Finley JD, Hofstadter MB. The Yale Preoperative Anxiety Scale: how does it compare with a "gold standard"? *Anesth Analg.* 1997;85:783-788.
74. Laufenberg-Feldmann R, Kappis B. Assessing preoperative anxiety using a questionnaire and clinical rating: a prospective observational study. *Eur J Anaesthesiol.* 2013;30:758-763.
75. Jansson-Frojmark M, Harvey AG, Lundh LG, Norell-Clarke A, Linton SJ. Psychometric properties of an insomnia-specific measure of worry: the Anxiety and Preoccupation about Sleep Questionnaire. *Cogn Behav Ther.* 2011;40:65-76.
76. van den Bergh RC, Korfage IJ, Borsboom GJ, Steyerberg EW, Essink-Bot ML. Prostate cancer-specific anxiety in Dutch patients on active surveillance: validation of the memorial anxiety scale for prostate cancer. *Qual Life Res.* 2009;18:1061-1066.
77. De Peuter S, Janssens T, Van Diest I, et al. Dyspnea-related anxiety: the Dutch version of the Breathlessness Beliefs Questionnaire. *Chron Respir Dis.* 2011;8:11-19.
78. Migliore NA, Whiteson J, Demetis S, Rey M. A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire. *J Cardiopulm Rehabil.* 2006;26:395-404.
79. Kuhl K, Kuhn C, Kenn K, Rief W. [The COPD-Anxiety-Questionnaire (CAF): a new instrument to assess illness specific anxiety in COPD patients]. *Psychother Psychosom Med Psychol.* 2011;61:e1-e9.
80. Carleton RN, Asmundson GJ. The multidimensionality of fear of pain: construct independence for the fear of Pain Questionnaire-Short Form and the Pain Anxiety Symptoms Scale-20. *J Pain.* 2009;10:29-37.
81. Mindham J, Espie CA. Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID): development and psychometric properties of a new measure for use with people with mild intellectual disability. *J Intellect Disabil Res.* 2003;47:22-30.
82. Ford J, Finch JF, Woodrow LK, et al. The Florida Shock Anxiety Scale (FSAS) for patients with implantable cardioverter defibrillators: testing factor structure, reliability, and validity of a previously established measure. *Pacing Clin Electrophysiol.* 2012;35:1146-1153.
83. McKinley S, Madronio C. Validity of the Faces Anxiety Scale for the assessment of state anxiety in intensive care patients not receiving mechanical ventilation. *J Psychosom Res.* 2008;64:503-507.
84. Court H, Greenland K, Margrain TH. Content development of the Optometric Patient Anxiety Scale. *Optom Vis Sci.* 2007;84:729-737.
85. Levin JS. The factor structure of the pregnancy anxiety scale. *J Health Soc Behav.* 1991;32:368-381.
86. Blin O, Azorin JM, Lecrubier Y, Souche A, Fondarai J. [The Psychotic Anxiety Scale (PAS). Evaluation of inter-rater reliability and correspondence factorial analysis]. *Encephale.* 1989;15:543-547.
87. Skolnick RB, Schare ML, Wyatt KP, Tillman MA. Aviophobia assessment: validating the Flight Anxiety Situations Questionnaire as a clinical identification measure. *J Anxiety Disord.* 2012;26:779-784.
88. Silove D, Manicavasagar V, O'Connell D, Blaszczyński A, Wagner R, Henry J. The development of the Separation Anxiety Symptom Inventory (SASI). *Aust N Z J Psychiatry.* 1993;27:477-488.
89. Gerolimatos LA, Gould CE, Edelman BA. Exploratory factor analysis of the Anxiety Control Questionnaire among older adults. *Behav Modif.* 2012;36:600-616.
90. Magnavita N. [Anxiety and depression at work. The A/D Goldberg Questionnaire]. *G Ital Med Lav Ergon.* 2007;29:670-671.
91. Hopko DR, Mahadevan R, Bare RL, Hunt MK. The Abbreviated Math Anxiety Scale (AMAS): construction, validity, and reliability. *Assessment.* 2003;10:178-182.
92. Vigil-Colet A, Lorenzo-Seva U, Condon L. Development and validation of the statistical anxiety scale. *Psicothema.* 2008;20:174-180.
93. Muschalla B, Linden M, Olbrich D. The relationship between job-anxiety and trait-anxiety—a differential diagnostic investigation with the Job-Anxiety-Scale and the State-Trait-Anxiety-Inventory. *J Anxiety Disord.* 2010;24:366-371.
94. Motl RW, Conroy DE. Validity and factorial invariance of the Social Physique Anxiety Scale. *Med Sci Sports Exerc.* 2000;32:1007-1017.
95. Momozawa Y, Terada M, Sato F, et al. Assessing equine anxiety-related parameters using an isolation test in combination with a questionnaire survey. *J Vet Med Sci.* 2007;69:945-950.
96. Cirakoglu OC, Senturk GC. Development of a performance anxiety scale for music students. *Med Probl Perform Art.* 2013;28:199-206.
97. Becker J. [Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT)]. *Computerized Adaptive Test (CAT) for Anxiety developed based on the Item Response Theory (IRT).* Berlin, Germany: FU Berlin, 2004.
98. Becker J, Fliege H, Kocalevent RD, et al. Functioning and validity of a Computerized Adaptive Test to measure anxiety (A-CAT). *Depress Anxiety.* 2008;25:E182-E194.
99. Walter O, Becker J, Fliege H, et al. Development of a Computer Adaptive Test for Anxiety (A-CAT). *Diagnostica.* 2005;51:88-100.
100. Rose M, Bjorner JB, Fischer F, et al. Computerized adaptive testing-ready for ambulatory monitoring? *Psychosom Med.* 2012;74:338-348.
101. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of the CAT-ANX: A Computerized Adaptive Test for Anxiety. *Am J Psychiatry.* 2014;171:187-194.
102. Wahl I, Lowe B, Bjorner JB, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol.* 2014;67:73-86.
103. Rose M, Hess V, Horhold M, Braehler E, Klapp BF. [Mobile computer-assisted psychometric diagnosis. Economic advantages and results on test stability]. *Psychother Psychosom Med Psychol.* 1999;49:202-207.
104. Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. *Qual Life Res.* 2009;18:125-136.
105. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol.* 2014;67:108-113.
106. Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res.* 2014;23:1609-1618.
107. Court H, Greenland K, Margrain TH. Measuring patient anxiety in primary care: Rasch analysis of the 6-item Spielberger State Anxiety Scale. *Value Health.* 2010;13:813-819.
108. Lowe B, Decker O, Muller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care.* 2008;46:266-274.
109. Zung WW. A rating instrument for anxiety disorders. *Psychosomatics.* 1971;12:371-379.
110. Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, McCormick RA. Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J Abnorm Psychol.* 1995;104:3-14.
111. Watson D, Clark LA, Weber K, Assenheimer JS, Strauss ME, McCormick RA. Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *J Abnorm Psychol.* 1995;104:15-25.
112. Meyer T, Miller M, Metzger R, Borkovec T. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther.* 1990;28:487-495.

113. Wittchen HU, Boyer P. Screening for anxiety disorders. Sensitivity and specificity of the Anxiety Screening Questionnaire (ASQ-15). *Br J Psychiatry Suppl.* 1998;10-17.
114. Norton PJ, Robinson CM. Development and evaluation of the anxiety disorder diagnostic questionnaire. *Cogn Behav Ther.* 2010;39:137-149.
115. Scheier IH, Cattell RB, Sullivan WP. Predicting anxiety from clinical systems of anxiety. *Psychiatr Q Suppl.* 1961;[suppl] 35:114-126.
116. Belleville G, Belanger L, Ladouceur R, Morin CM. [Sensitivity and specificity of the worry and anxiety questionnaire (WAQ) in a sample of health-care users in the province of Quebec]. *Encephale.* 2008;34:240-248.
117. Scholing A, Emmelkamp PM. Self report assessment of anxiety: a cross validation of the Lehrer Woolfolk Anxiety Symptom Questionnaire in three populations. *Behav Res Ther.* 1992;30:521-531.
118. Koksal F, Power KG. Four Systems Anxiety Questionnaire (FSAQ): a self-report measure of somatic, cognitive, behavioral, and feeling components. *J Pers Assess.* 1990;54:534-545.
119. Morris LW, Davis MA, Hutchings CH. Cognitive and emotional components of anxiety: literature review and a revised worry-emotionality scale. *J Educ Psychol.* 1981;73:541-555.
120. Lopez P, Jose A, Olivares Rodriguez J, Sanchez-Garcia R. Modelo de Rasch aplicado a la version espanola del autoinforme Liebowitz Social Anxiety Scale for Children and Adolescents (LSAS-CA-SR). *Int J Clin Health Psychol.* 2008;8:233-245.
121. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol.* 2007;46:1-18.
122. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry.* 2012;69:1104-1112.
123. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol.* 1959;32:50-55.
124. Fergus TA, Valentiner DP, McGrath PB, Gier-Lonsway SL, Kim HS. Short forms of the Social Interaction Anxiety Scale and the Social Phobia Scale. *J Pers Assess.* 2012;94:310-320.
125. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther.* 1998;36:455-470.
126. Safren SA, Turk CL, Heimberg RG. Factor structure of the Social Interaction Anxiety Scale and the Social Phobia Scale. *Behav Res Ther.* 1998;36:443-453.
127. Connor KM, Davidson JR, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. *Br J Psychiatry.* 2000;176:379-386.
128. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JR. Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. *Depress Anxiety.* 2001;14:137-140.
129. Caballo VE, Salazar IC, Iruiria MJ, Arias B, Hofmann SG. The multidimensional nature and multicultural validity of a new measure of social anxiety: the Social Anxiety Questionnaire for Adults. *Behav Ther.* 2012;43:313-328.
130. Rytwinski NK, Fresco DM, Heimberg RG, et al. Screening for social anxiety disorder with the self-report version of the Liebowitz Social Anxiety Scale. *Depress Anxiety.* 2009;26:34-38.
131. Safren SA, Heimberg RG, Horner KJ, Juster HR, Schneier FR, Liebowitz MR. Factor structure of social fears: The Liebowitz Social Anxiety Scale. *J Anxiety Disord.* 1999;13:253-270.
132. Gibbons RD, Weiss DJ, Kupfer DJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv.* 2008;59:361-368.