



Original article

Revised manual for the Global Assessment of Functioning scale

G. Pedersen^{a,b,*}, Ø. Urnes^a, B. Hummelen^c, T. Wilberg^{c,d}, E.H. Kvarstein^{a,d}^aOslo University Hospital, Department of Personality Psychiatry, Division of Mental Health and Addiction, Oslo, Norway^bNORMENT, KG Jebsen Center for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Norway^cDepartment of Research and Development, Division of Mental Health and Addiction, Oslo University Hospital, Norway^dInstitute of Clinical Medicine, University of Oslo, Norway

ARTICLE INFO

Article history:

Received 26 September 2017

Received in revised form 22 December 2017

Accepted 30 December 2017

Available online 3 March 2018

Keywords:

Assessment

GAF

Global assessment of functioning

ABSTRACT

Global Assessment of Functioning (GAF) is a single measure of overall psychosocial impairment caused by mental factors, constituting Axis V of the Diagnostic and Statistical manual of Mental disorders, third and fourth versions. Despite its widespread use, several challenges and shortcomings have been discussed the last three decades. The current article describes some of the more serious challenges of the GAF manual, and presents a revised version more in accordance with the nature of this clinical construct. Some crucial aspects of the understanding of GAF and general guidelines for scoring are also discussed.

© 2018 Elsevier Masson SAS. All rights reserved.

1. Background

The first standardised instrument for assessing patients overall mental health was introduced more than 50 years ago when Lester Luborsky introduced his Health-Sickness Rating Scale (HSRS) [1]. Later on, Endicott and colleagues [2] modified the original instrument into the Global Assessment Scale (GAS). Both HSRS and GAS are single 100-point rating scales reflecting overall functioning, and are evaluated on a continuum ranging from score 1, representing the hypothetically sickest, to score 100, representing the hypothetically healthiest individual. In the third edition of the Diagnostic and Statistical Manual of Mental disorders (DSM-III) [3] axis V was introduced as a measure of “adaptive functioning”, scored on a 7-point scale ranging from superior to grossly impaired. In DSM-III-R [4], the Global Assessment of Functioning scale (GAF) replaced this axis V for an observer-based assessment of psychological, social, and occupational functioning. The GAF scale was based on the GAS, although the upper range from 91 to 100 was omitted. Within DSM-IV [5], the GAF scale was extended to a 100-point scale. In 2013, the American Psychiatric Association [6] introduced the fifth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). In this version, the axis system was removed, and GAF is no longer an element of the DSM classification system [7].

The GAF is intended to be a single measure of overall impairment caused by mental factors. It's intended use is to communicate the level of severity and impairment, indicate the need of professional help, and reflect improvement or change over time. It is a generic measure, not related to any specific diagnosis. By reflecting the level of severity, GAF provides important additional information to the categorical diagnostic classifications, and the extensive use of this measure over the years confirms its importance.

The reliability of GAF scores has been proven to be acceptable, especially under conditions when raters are experienced and trained [8–13]. As to the validity of GAF, several studies have focused on its associations with other clinical phenomena. Some of these studies have found the validity to be only modest [14–16], but other studies have found significant associations between GAF scores and the presence of axis-II pathology, self reported symptom distress, interpersonal problems, as well as social functioning [9,17–18].

However, there has been, and still is, some scepticism concerning the use of one single scale to assess both the level of psychological symptoms and social and occupational functioning [14,19–22]. Since these three dimensions do not always vary together it will constitute a challenge to both the reliability of the GAF scores and the validity of the inferences drawn from these. This scepticism has led to several studies of the GAF scale, and a review by Goldman and colleagues [14] suggested the need for different measures for different areas of functioning. As a potential solution, they introduced the ‘Social and Occupational Functioning Assessment Scale’ (SOFAS), which became a supplement for further research under axis V of the DSM-IV. However, in routine

* Corresponding author at: Department of Personality Psychiatry, Division of Mental Health and Addiction, Oslo University Hospital, Ullevaal, PO Box 4956 Nydalen, 0424 Oslo, Norway.

E-mail address: geir.pedersen@medisin.uio.no (G. Pedersen).

clinical settings the traditional GAF scale is far more frequently used than SOFAS.

In the GAF manual, ten main intervals are described by examples of symptoms and functional impairment, separated by “and/or” in the text. This makes it easy to form two separate scales; one for symptoms- and one for functional impairment. A study of such a simple split version of the traditional GAF scale was conducted by Jones and colleagues [23], concluding that these two separate scales had different patterns of validity. Then, in 1998 Karterud and colleagues [24] constructed a Norwegian version of a split GAF scale. The manual was the same as for the original GAF scale except that the symptom- and function descriptions were kept on different sheets and rated separately. The validity of this split version is reported in a study by Pedersen and Karterud [18].

In 1980 WHO published the International Classification of Impairments, Disabilities and Handicaps (ICIDH). After some years of revision, the International Classification of Functioning, Disability and Health (ICF) was introduced, and later operationalized by the WHO Disability Assessment Schedule (WHODAS) [25]. In several ways ICF is a reasonable alternative to GAF since it represents a generic measure covering six different areas of disability: Cognition (understanding and communicating), Mobility (getting around), Self-care, Interpersonal (getting along), Life activities (Household and Work), and Participation in society. However, as it is more comprehensive than GAF, a full interview based assessment is more time consuming. Another reason to assume that WHODAS would not completely replace GAF as a routine measure among mental health care professionals and researchers is that WHODAS is merely a measure of functioning, and not focused on the level of symptom severity. In this respect WHODAS has some clear disadvantages compared to GAF as a routine measure of change with respect to overall levels of severity.

2. Challenges and shortcomings of GAF

There is still a potential for improvement of the GAF manual as we know it today [20,26]. A qualitative study emphasized several weaknesses of the GAF manual [26]. Clinicians reported that poor examples and lack of continuity caused difficulties in choosing an interval or distinguishing between intervals. Moreover, essential misunderstandings of the nature of GAF were also disclosed. Some raters preferred specific, clear, and unambiguous criteria, and examples, and some wanted rules of thumbs – like that a GAF score of 40 was a cut-off level for psychotic symptoms.

Of specific concern are several ambiguous keywords and anchor points along the scale, which gives an unfortunate width of interpretations, introducing decreased levels of reliability and clinical utility. Of most concern is the lack of continuity, especially at the low end of the scale where GAF goes from addressing general severity to danger of hurting oneself or others (GAF levels 1–20). A classical example of the last was seen in the report from the first court-appointed psychiatrists in the Norwegian mass murder case (The “Breivik-case”), where extremely low GAF scores were given [27–29]. Here, the psychiatrists were right with respect to the GAF manual, but wrong according to GAF as a continuous clinical construct. Danger of hurting oneself or others is certainly present among people with scores higher up on the GAF scale. With keywords such as recurrent violence and suicide attempts, the GAF manual gives examples of behavior or symptoms that have nothing to do with GAF. GAF is not a measure of suicide risk or violence. As it is, the GAF manual does not represent a continuous hypothetical clinical construct.

The GAF manual instructs raters to consider psychological, social, and occupational functioning on a hypothetical

continuum of mental health-/illness. This implies that specific symptoms or impairments in social functioning have to be evaluated at a more general level of clinical relevance, and not isolated. Moreover, each of the ten intervals of the GAF manual should represent a gradual increase in symptom severity and functional impairment. The keywords and examples within each of the intervals have to be descriptive, delineating, and clinically relevant. According to this, obvious weakness was apparent in the current manual. Some examples were not descriptive in respect to what GAF is measuring, i.e. ‘frequent shoplifting’, ‘occasional truancy’, ‘theft within the household’. Some are less delineating, such as ‘occasional argument with family members’, ‘few friends’, ‘no friends’, and ‘acts grossly inappropriately’, or less relevant, such as ‘child frequently beats up younger children, defiant at home’.

3. A moderate, but necessary revision of the GAF manual

After years of dealing with and explaining these shortcomings during teaching courses in GAF assessment, a group of experienced researchers and clinicians from the Research Group for Personality Psychiatry at the University of Oslo, and The Norwegian National Advisory Unit for Personality Psychiatry, at the Oslo University Hospital, Norway, recently, decided to improve the most prominent shortcomings of the GAF manual. This resulted in a new, revised split version of GAF (Appendix A in Supplementary material), where some keywords and examples were changed and/or replaced, so that the intervals represent a clinically coherent continuum. This newly, revised GAF scale does not represent a radical change or substantial shift in the meaning of the construct. For raters who have scored GAF with an overall understanding of its continuity, the revision represents little, if any, practical importance. The most prominent changes are among the descriptions and examples at GAF levels 1–30, and especially between 1 and 20. The revised GAF manual has replaced unfortunate phrases and examples with new, more descriptive, relevant and unambiguous examples. Furthermore, the pedagogic intervals are made more coherent with respect to the dimensionality of the GAF construct.

4. The nature of GAF and its implications for general scoring guidelines

GAF is a clinical construct without any natural measurement scale or unit of measurement. It is purely something we imagine, and it is crucial that raters of GAF share the same imagination. A common understanding of the GAF construct has therefore to be established. Furthermore, common scoring instructions must be made. Moreover, reliable and valid GAF scores depend not only on the raters obtaining understanding by reading manuals or instructions, but at least as important, on practice, clinical experience, and on calibrating one’s ratings by discussing with colleagues.

As it is constructed, the GAF scale represents a latent continuous measure scored from 1 to 100, where score 1 represents the worst imaginable level of symptom severity and impairment of psychosocial functioning, and score 100 represents the, hypothetically, most optimal level. For such a hypothetical scale to capture all observable variations, it is implicit that the top (100) and bottom (1) of the scale are merely theoretical sizes. The GAF scale includes ten intervals, but not distinct cut-off levels. The examples and keywords within each interval should be seen as a description of the midist point within the interval, i.e. 45, 55, 65, not a description of the whole interval. The evaluations should then focus on whether a patient’s score is at, above, or below this score-level.

The GAF manual should be seen as an aid for deciding on GAF scores in a comprehensive way. The given symptoms or examples within each interval are neither sufficient nor exhaustive, but are selected for illustration – there may be many other equally representative examples. Together they represent a continuum. If focusing too strongly on the exact given examples, an overall perspective will be lost. As an example, a generally, moderately well-functioning person with a GAF score of 67 might have had a single episode of psychotic experience. If missing this point, the reliability of GAF scores will decrease.

The GAF focuses on assessment of severity. The single GAF score represents both the most severe level of symptom or function which is reported, and at the same time, the level most representative of the person, situation, and period in question. In some cases, it may be important not to overemphasize single episodes. Symptoms and psychosocial functioning may vary from week to week or from one day to another. Examples are occasional panic attacks, anxiety that comes and goes, better days within a state of moderate depression, or feelings of hopelessness and suicide thoughts lasting some hours. In such cases, the clinical judgment of the rater is crucial. The GAF score signalizes the severity of a current, overall situation.

Assessment of the symptom level of GAF is focused on the severity of the symptoms, extent, duration, and the consequences this has for the person's self-perception and quality of life. Assessment of impairment of functioning is focused on the person's actual role functions, such as student, employee, romantic partner, spouse, parent, friend, neighbor etc. This simply means the ability to set and reach one's goals in life, care for one self and others and maintain necessary daily functioning. However, the level of functioning has two aspects; a quantitative aspect (frequencies of physical exercise, visiting cinema with friends, going to work, keeping contact with family) and a qualitative aspect (quality of the relations, involvements and personal engagement in activities etc.). Both aspects have to be considered. Despite high levels of activity, the function score will be reduced if the quality and personal engagement is low or absent.

Within the original GAF manual suicide thoughts and attempts immediately caused low scores. In the present revised manual, suicide issues are implicit in the severity level of the symptom score. The same reasoning applies to danger of hurting others. Thus, it is the underlying symptom level that is relevant. As noted, GAF is not a measure of suicide risk or violence. In general, there is no specific behavior that is related to any score or cut-point of the GAF scale.

Revisions of measurement methods or instruments, as the current revised GAF manual, warrants further studies focusing on reliability and validity. Currently, a study of reliability comparing estimates from the original and revised GAF manual is performed at the Oslo University Hospital, Norway, approved by the Norwegian Regional Committee for Medical Research and Ethics, ref. no.: 2015/359).

5. Conclusion

In the current revised GAF manual, unfortunate phrases and examples are replaced with new, more descriptive, relevant and unambiguous examples. Furthermore, pedagogic intervals are made more coherent with respect to the dimensionality of the GAF construct. However, scoring guidelines and teaching courses in GAF should focus more on the general continuity of this clinical construct rather than on the specific examples of symptoms and functional impairment listed in the manual. As noted, the GAF manual is not a definition of the construct to be taken literally. With this in mind, and with the current revised split version of the GAF manual, courses in GAF scoring may be more efficient and

focused, discussions and arguments concerning specific scores should be less tempered and confusing, and lastly, GAF ratings should become more reliable, leading to improved validity of the inferences drawn from single scores. More subsequent studies related to reliability and validity of the revised GAF manual are needed.

Conflict of interest

None of the above authors have any financial disclosure/conflict of interest related to this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.eurpsy.2017.12.028>.

References

- [1] Luborsky L. Clinicians' judgements of mental health. *Arch Gen Psychiatry* 1962;7:407–17.
- [2] Endicott J, Spitzer RL, Fleiss JF, Cohen J. The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 1976;33:766–71.
- [3] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd edn Washington, DC: American Psychiatric Association; 1981.
- [4] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd edn Washington, DC: American Psychiatric Association; 1987.
- [5] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edn Washington, DC: American Psychiatric Association; 1994.
- [6] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edn Arlington, VA: American Psychiatric Publishing; 2013.
- [7] American Psychiatric Association. Insurance Implications of DSM-5. 2013 https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM_Insurance-Implications-of-DSM-5.pdf (26.9.2017).
- [8] Dworkin RJ, Friedman LC, Telschow RL, Grant KD, Moffic HS, Sloan VJ. The longitudinal use of the Global Assessment Scale in multiple-rater situations. *Community Ment Health J* 1990;26:335–44.
- [9] Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR, et al. Reliability and validity of DSM-IV axis V. *Am J Psychiatry* 2000;157:1858–63.
- [10] Løv Dahl H, Friis S. Routine evaluation of mental health: reliable information or worthless 'guesstimates'? *Acta Psychiatrica Scandinavica* 1996;93:125–8.
- [11] Pedersen G, Hagtvet KA, Karterud S. Generalizability studies of the Global assessment of functioning (GAF) – Split version. *Compr Psychiatry* 2007;48:88–94.
- [12] Startup M, Jackson MC, Bendix S. The concurrent validity of the global assessment of functioning (GAF). *Br J Clin Psychol* 2002;41:417–22.
- [13] Vatnaland T, Vatnaland J, Friis S, Opjordsmoen S. Are GAF scores reliable in routine clinical use? *Acta Psychiatrica Scandinavica* 2007;115:257–336.
- [14] Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: A review of measures of social functioning. *Am J Psychiatry* 1992;149:1148–56.
- [15] Moos RH, Nichol AC, Moos BS. Global assessment of functioning ratings and the allocation and outcomes of mental health services. *Psychiatric Serv* 2002;53:730–7.
- [16] Roy-Byrne P, Dagadakis C, Unutzer J, Ries R. Evidence for limited validity of the revised global assessment of functioning scale. *Psychiatric Serv* 1996;47:864–6.
- [17] Karterud S, Pedersen G, Bjordal E, Brabrand J, Friis S, Haaseth Ø, et al. Day hospital treatment of patients with personality disorders. Experiences from a Norwegian treatment research network. *J Personal Disord* 2003;17:173–93.
- [18] Pedersen G, Karterud S. The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. *Compr Psychiatry* 2012;53:292–8.
- [19] Bacon SF, Collins M, Plake EV. Does the global assessment of functioning assess functioning. *J Mental Health Counsel* 2002;24:202–12.
- [20] Monrad Aas IH. Global Assessment of Functioning (GAF): properties and frontier of current knowledge. *Ann Gen Psychiatry* 2010. doi:<http://dx.doi.org/10.1186/1744-859X-9-20>.
- [21] Schwartz RC, Del Prete-Brown T. Construct validity of the Global Assessment of Functioning Scale for clients with anxiety disorder. *Psychol Rep* 2003;92:548–50.
- [22] Skodol AE, Link BG, Shrout PE, Horwath E. The revision of axis V in DSM-III-R: should symptoms have been included? *Am J Psychiatry* 1988;145:825–9.
- [23] Jones SH, Thornicroft G, Coffey M, Dunn G. A brief mental health outcome scale-reliability and validity of the Global Assessment of Functioning (GAF). *Br J Psychiatry* 1995;166:654–9.
- [24] Karterud S, Pedersen G, Løv Dahl H, Friis S. Global assessment of functioning – split version. Background and scoring guidelines. Oslo: Dep. of Psychiatry, Ullevaal University Hospital; 1998.
- [25] Üstün TB, Kostanjsek N, Chatterji S, Rehm J, editors. Measuring Health and Disability. Manual for WHO Disability Assessment Schedule (WHODAS 2.0). Geneva, Switzerland: World Health Organization; 2010.

- [26] Monrad Aas IH, Sonesson O, Torp S. A qualitative study of clinicians experience with rating of the global assessment of functioning (GAF) scale. *Commun Ment Health J* 2016, doi:<http://dx.doi.org/10.1007/s10597-016-0067-6>.
- [27] Foss AB, Johansen PA, Andreassen TA. Breivik fikk 2 av 100 mulige poeng i psykiatrisk test. *Aftenposten*. 2011 03.12.2011. <http://www.aftenposten.no/nyheter/iriks/Breivik-fikk-2-av-100-mulige-poeng-i-psykiatrisk-test-6712735.html#.UzF23KIVAs0> (28.6.2017).
- [28] Gardell M. Galna slutsatser. *Aftonbladet*. 2011 07.12.2011. <http://www.aftonbladet.se/kultur/article14042903.ab> (28.6.2017).
- [29] Kissane K. Mad or bad? The jury is out. *The Sydney Morning Herald*. 2012 14.04.2012, <http://www.smh.com.au/world/mad-or-bad-the-jury-is-out-20120413-1wyvg.html> (28.6.2017)..