Towards a Better Global Assessment of Functioning (GAF): Improving Scale Properties

IH Monrad Aas

Research Unit, Division of Mental Health and Addiction, Vestfold Hospital Trust, Tonsberg, Norway

ABSTRACT: Background: *Global Assessment of Functioning (GAF) is a rating scale used in a very high number of studies. GAF is known worldwide. GAF rates severity of illness in psychiatry and is often used together with instruments rating other characteristics of mental disease. From research, we know there are problems with GAF (for example, reliability and validity problems). The properties of the GAF scale need a closer examination with the potential for improvement in mind. The present study has focus on GAF properties.* **Purpose:** *to show both gaps in current knowledge and ideas for further development.* **Methods:** *The present study is based upon a systematic literature review.* **Findings:** *for the properties of GAF, numerous gaps in knowledge were found: for example, a continuous scale is used for the present GAF, but would a categorical scale make a better GAF? On visual scales scoring is done by setting a mark directly on the scale, but would transformation to a visual scale result in an improved GAF? The anchor points (including examples) were decided early in the history of GAF, but would new anchor points and examples result in a better GAF (anchor points for symptoms, functioning, positive mental health, prognosis, improvement of generic properties, exclusion criteria for scoring in each 10-point intervals, and anchor points at the endpoints of the scale)? Is a change in the number of anchor points and their distribution over the total scale important? Rating within 10-point intervals can be requiring, but can better instructions improve this? Internationally, GAF with both one and two values are used, but what is the advantage of having separate symptom (GAF-S) and functioning scales (GAF-F)? GAF-S and GAF-F scales should score different dimensions and still be correlated, but what is the best combination of definitions for GAF-S and GAF-F?* **Conclusions:** *Given the widespread use, research-based development of GAF has not been especially strong. Further research could improve GAF.*

Key words: *GAF, Global Assessment of Functioning, psychiatry, methodology*

INTRODUCTION

A large number of scoring systems are developed for psychiatry. The Global Assessment of Functioning (GAF) is known worldwide, translated into many languages, and used in many outcome studies (Aas, 2010; 2011; 2014). GAF is used to rate severity of illness in psychiatry and covers the range from positive mental health to severe psychopathology. It is an overall (global) measure of how patients are doing (Moos et al., 2000; Rosse & Deutsch, 2000). GAF is not intended to be a diagnosis-specific scoring system, but a generic. Compared to diagnosis, GAF values represent more multidimensional information (Rosse & Deutsch, 2000; Schorre & Vandvik, 2004). The degree of mental illness is measured by rating psychological, social and occupational functioning (Goldman et al., 1992; Vatnaland et al., 2007). The simplicity of GAF is an advantage (Aas, 2010).

Internationally, recording GAF is either done with a single value (this is the most severe of the symptom and functioning values) or both symptom (GAF-S) and functioning (GAF-F) values are recorded. The symptom and functioning scales have both 100 scoring possibilities (1-100). The 100-point scales are divided into 10 intervals, or sections, each with 10 scoring possibilities (examples: 31-40 and 51-60). Verbal instructions (called anchor points) describe symptoms and functioning relevant for scoring in the 10-point intervals. The anchor points represent hierarchies of mental illness (McDowell & Newell, 1987; Pedersen et al., 2007; Vatnaland et al., 2007). The

anchor points for interval 1-10 describe the most severely ill and the anchor points for interval 91-100 describe the healthiest. In addition to anchor points, examples are found for each 10-point interval. The examples are intended to help with the scoring in each interval. For example, in the interval 51-60 (moderate symptoms) on the symptom scale, patients with occasional panic attacks can be rated, and in the interval 51-60 (moderate difficulty in social, occupational or school functioning) on the functioning scale, patients with conflicts with peers or co-workers and few friends can be rated (Karterud et al., 1998; Schorre &Vandvik, 2004). The finer grading within intervals (for example 32, 35, 37 and 55, 57, 59) provides the possibility of distinguishing between nuances (Thomson, 1989), but there are no verbal instructions for this grading found on either the two scales. Research on GAF shows problems with both reliability and validity. Reliability studies show the extreme 20% of raters to account for more than 50% of the spread of scores and deviations can be 20 points or more (Loevdahl & Friis, 1996; Vatnaland et al., 2007). Different studies show inter-rater reliability to be highly variable, but it should be noticed that this includes very good reliability. Reliability seems to be lower in routine clinical practice than in research (Burlingame et al., 2005; Hilsenroth et al., 2000; Moos et al., 2000; Soderberg et al., 2005; Startup et al., 2002; Vatnaland et al., 2007). Concurrent validity (Bates et al., 2002; Burlingame et al., 2005; Goldman et al., 1992; Hall, 1995; Hay et al., 2003; Hilsenroth et al., 2000; Jones et al., 1995; Niv et al., 2007; Patterson & Lee, 1995; Pedersen et al., 2007; Piersma & Boes , 1997; Robert et al., 1991; Roy-Byrne et al., 1996; Salvi et al. 2005; Tungstrom et al., 2005) and predictive validity (Bacon et al., 2002; Fallmyr & Repal, 2002; Goldman et al.,

*Correspondence regarding this article should be directed to: monrad.aas@siv.no

1992; Hay et al., 2003; Moos et al., 2000; Niv et al., 2007; Parker et al., 2002) are more problematic. There are few empirical results for GAF sensitivity (Bird et al., 1987).

In the clinic, the primary goal of the assessment process is to contribute to the solution of a person's problems (Bruyn, 2003). A generic and global scoring system, such as GAF, that covers the range from positive mental health to severe psychopathology has advantages for clinical practice (for example, routine quality assessment of treatment, supplementing scales that give more detail) (Lingjaerde et al., 1989), research (for example, comparison of treatment outcome across diagnoses), and policy and management planning (for example, allocation of resources, measurement of case-mix in psychiatric organizations). We are dealing with a wide range of potential applications and GAF must be good enough for its purposes. To dismiss an existing instrument due to problems can be a too simple solution (Streiner & Norman, 1994). Work to improve GAF is an alternative. Further development for GAF means work to improve validity and reliability, and to ensure good sensitivity, and generic properties.

The present study is based upon the first of three systematic literature reviews (Aas, 2010; 2011; 2014). The purpose of the study is to show the gaps in current knowledge, and ideas about further development when it comes to properties of the GAF scale.

PROPERTIES OF GAF AND GAPS IN KNOWLEDGE

Properties of GAF are defined as characteristic traits or attributes that serve to define GAF (or may have a role to define a future new GAF). The gaps identified in the present study are defined as properties of GAF where no, or little, research has been done, with characteristics that suggest further development is likely to have a role for improvement of GAF.

The first of the three systematic literature reviews (Aas, 2010) shows the properties of the GAF scale in four main categories. These main categories (including subcategories) are important when it comes to further development of GAF and further development means work to improve GAF. The four main categories are: (1) scaling; (2) the anchor points of GAF; (3) scoring within 10-point intervals; and (4) the number of scales.

Scaling

For science in general, measurement and scaling are fundamental, but not less important for evaluation of interventions in health care. Problems with quantification play a key role for the reliability of health care interventions. Scaling means quantifying qualities by assigning numbers (Young, 1984). For the future development of psychiatry, scaling will be important (Bech et al., 1993; Breakwell & Millard, 1995; McDowell & Newell, 1987; Nunally & Bernstein, 1994; Widiger & Clark, 2000).

Continuous or Categorical Scale

Continuous and categorical (i.e. discrete) scales are two different scale types. In GAF, a finely graded continuous scale (graded with 100 scoring possibilities) has been preferred to a categorical scale. Classification into categories, with verbally formulated inclusion criteria for each category, is an alternative to continuous scales.

Gap in knowledge

The development of GAF has little basis in general research on what is best for a global functioning scale, i.e. a continuous or categorical scale. Little research has been done on GAF concerning whether a continuous or categorical scale is better.

Visual Scale

A straight line with anchor points at each end, to indicate the extremes, is called a visual analogue scale or a VAS. The severity of the phenomenon is scored by marking a point on the scale. The scored value is found by measuring the distance from the scale's lower end to the point.

Gap in knowledge

A VAS could be an alternative to the present GAF, but we do not know whether scoring by marking a point on the scale on improves scoring. The VAS could be equipped with anchor points along the line, but we do not know if the present GAF's anchor points are the best, if the number of anchor points should be changed, and if change in location along the line for the anchor points will be right.

Scales and Further Treatment of Data

In some research projects, collected raw data for GAF are merged into a limited number of categories (Moos et al., 2000; Moos et al., 2002). Merging functioning into just two categories ('superior to fair' and 'poor to grossly impaired') is known in the literature and is the simplest categorization (Schrader et al., 1986). However, when statistical analysis is done such dichotomization may well give different conclusions than a finer grading with an average of raw data GAF values. For a single scale GAF, 'whichever is the worse' of an individual's symptom and functioning values is the GAF score (First, 1995). As the symptom and functioning scales of GAF do not score the same, recording just one figure can be criticised. We are dealing with an obvious loss of information.

Gap in knowledge

Merging raw data into two (or a few) categories may well make conclusions from statistical analysis vulnerable to error, but the issue has been given little attention and is little analyzed. The practise in the single scale GAF of recording just one score has not been subjected to, much scrutiny.

The Anchor Points of GAF

In psychiatry, severity of illness is often expressed with symptoms and functioning, but other factors have a role to play. Different psychiatric diagnoses express differences in severity. Likewise, stage of development of the illness, intensity (for example frequency and duration of periods with symptoms over a time period), and co-morbidity (Aas, 1991; Seligman & Csikszentmihayi, 2000; Seligman et al., 2005; Wells et al., 1989).

The Nature of Anchor Points

The 10 anchor points, with examples of symptoms and functioning items, give a general idea on what to stress in scoring GAF. The use of examples is important and is likely to improve assessment (Rogers, 2001). Items used in different symptom and functioning scoring systems are different.

Gap in knowledge

GAF has a history half a century back in time, but the character of anchor points and examples is much the same as in the early version. As experimentation with other anchor points (and other examples) has hardly been done, we do not know if such change would result in an improved GAF. It is thinkable that other anchor points and examples would improve the generic properties, but which changes would lead to improvement is unknown. It is thinkable that other expressions of severity (like, stage of development of the illness, intensity, co-morbidity) could play a role for improvement, but we do not know. It is not self evident that all the rankings of the anchor points are correct, but this is hardly studied. Studies of reliability and validity exist, but comparison of reliability and validity for high and low values are difficult to find.

Symptoms

Much symptom research is performed since the early GAF

versions, but this has not at all resulted in a following change in GAF symptom anchor points. The symptom anchor points of today are much the same as those of the early versions. Symptom checklists are well known in today's psychiatry and can include questions about behavioural and somatic symptoms, and positive and negative feelings of well-being (McDowell & Newell, 1987; Sederer et al., 1995). When patients are asked about both positive feelings of well-being and somatic symptoms, the checklist is more objective and the intent of the measurement is concealed. Both sensitivity and specificity can be good (McDowell & Newell, 1987). Patients can have more than one symptom and the symptoms can be of different types and degrees of development. Symptoms can occur in clusters and clusters can be evaluated as basis for assessment of illness severity. Many symptoms in psychiatry have two aspects: form (e.g. auditory hallucination) and content (e.g. the person is told to do something) (Gelder et al., 2006).

Gap in knowledge

For the development of GAF, the considerable symptom research since GAF's early versions has played a minimal role. Learning from symptom research could play a role in work to improve symptom anchor points with examples. Analysis of symptom clusters, with different degrees of severity for each symptom, has not played much of a role for improvement of GAF scoring. Symptom content as a criterion for scoring illness severity is little studied.

Functioning

In the literature, we find many methods for rating of functioning (Aas, 2010; Bowling, 1993; Feinstein et al., 1986; Goldman et al., 1992; McDowell & Newell, 1987). A definition of functional status is: the degree to which an individual is able to perform socially allocated roles free of mentally (or physically) related limitations (Bowling, 1993). To develop a method for rating of functioning we need to decide:

• which type of functioning should be scored – to obtain a good image of overall functioning, it is necessary to rate several types of functioning, for example difficulties with participation in working life, daily activities, and social relationships

• different types of functioning can be graded in different ways and we need to decide how to grade each type

• whether an aggregate measure can be made, i.e. the total score expressed with one figure.

Gap in knowledge

For further development of GAF, the considerable international research on functioning has played a limited role. It is possible that the anchor points for functioning with their examples, and scoring within 10-point intervals could be improved by learning from research on functioning.

Positive Mental Health

In psychiatric research, focus on positive mental health factors has played a clearly smaller role than focus on illness itself (Seligman & Csikszentmihalyi, 2000; Vaillant, 2003). It is too simple to believe that positive and negative feelings are opposite ends of a single-dimension scale (McDowell & Newell, 1987). Positive health factors (such as life satisfaction, positive quality of life, psychological well-being, and even physical fitness) could be considered as factors of importance for scoring of GAF, but it is not much discussed (Bowling, 1993; Seligman & Csikszentmihalyi, 2000; Seligman et al., 2005).

Gap in knowledge

Work for identification of relevant positive mental health factors could play a role for the further development of GAF. Maybe use of positive health factors will improve the choice of 10-point interval, and the scoring within 10-point intervals.

Prognosis

The present GAF has limited value for assessing prognosis (Moos et al., 2002), and other systems predict prognosis better (Bowling, 1997; Burlingame et al., 2005; Parker et al., 2002). Prognosis is definable as a part of the severity of illness. A patient who is severely ill with a good prognosis can then be scored more highly than a patient who is less severely ill with a poor prognosis. Prognosis can be related to the patient's resources and not just the patient's problems (Bowling, 1993; Moos et al., 2000; Seligman & Csikszentmihalyi, 2000; Seligman et al., 2005).

Gap in knowledge

For GAF scoring, prognosis may be considered as a criterion, but this has not been given much attention. In the further development of GAF, study should be done of the importance of prognosis for scoring.

Exclusion Criteria

The 10-point intervals are defined by the anchor points. For rating in the 10-point intervals, the anchor points are inclusion criteria. Little work has been done to identify exclusion criteria for scoring in each interval.

Gap in knowledge

It is too simple to believe that exclusion criteria for scoring in each interval can be formulated as just the opposite of inclusion criteria. Little work has been done to identify exclusion criteria for scoring in each interval. Search for specific exclusion criteria could be a part of the future study of GAF.

Extremes of the GAF

In GAF, we find a hierarchy for severity of mental illness. The lowest and highest anchor points define the highest and lowest levels of severity. Endpoints show the variation possibilities for the severity phenomenon, and endpoints can influence scoring (Sutherland et al., 1983). In scoring of morbidity, perfect health often marks one extreme. In GAF-S, the other extreme is persistent danger of severely hurting self or others, and in GAF-F it is persistent inability to maintain minimal personal hygiene. Stages of disease are expressions of severity and disease-staging systems exist. In such a system, for a number of psychiatric conditions, death was used as the lowest rating possibility (Gonella, 1983).

Gap in knowledge

Study of the effect on GAF scores of different endpoints is something we know little about.

Number of Anchor Points

As the symptom and functioning scales both have 100 scoring possibilities, we are dealing with a good possibility to distinguish between nuances. However, verbal instructions are not found with a corresponding detail. There is a conflict between the high number of scoring possibilities and the limited verbal instructions. Having less than 100 scoring possibilities is not the only solution, but work to improve verbal instructions can be proposed, i.e. analysis aiming at adding more anchor points with examples (Pedersen et al., 2007). In work with psychiatric patients, the middle range of GAF is much used, and work can be done with a further detailing of anchor points and examples (Rey et al., 1995). For patients newly admitted to psychiatric treatment, scores for low severity are less frequent, and the solution can be to extend the verbal instructions used for the more severely ill (Endicott et al., 1976). Community studies can include people without need for treatment, and the question of

having extended instructions for the upper part of the scale can be raised.

Gap in knowledge

Systematic testing of different changes in the number of anchor points with examples, and their distribution over the total scale, to obtain a better GAF is difficult to find in the history of GAF.

Scoring within 10-Point Intervals

Endicott et al 1976 (Endicott et al., 1976) and the manual for DSM-IV-TR give instructions for scoring within 10-point intervals, but instructions are limited (Aas, 2011).

Gap in knowledge

Systematic study to improve scoring within 10-point intervals is limited. Categorical scales could be evaluated for the purpose. Such application of categorical scaling would require consideration of the nature and number of categories.

The Number of Scales

In the DSM-IV-TR instructions, raters are told to record only one figure for GAF, but both symptoms and functioning should be evaluated. The problem with recording only one figure is a resulting lack of knowledge if the figure is a functioning or symptom score.

GAF with two scales

In psychiatry, symptoms and functioning are often closely related (Goldman et al., 1992; Hilsenroth et al., 2000; Moos et al., 2000; Moos et al., 2002), but have been proposed to deviate frequently enough to recommend measuring both in outcome studies (Bacon et al., 2002; Goldman et al., 1992). GAF-S and GAF-F can be correlated with r= 0.61 (Pedersen et al., 2007).

Gap in knowledge

Symptoms and functioning are different dimensions, but knowledge about the advantage using GAF-S and GAF-F separately is limited. GAF-S and GAF-F score different dimensions, but the scores should still be correlated. Search for the right combination of definitions of GAF-S and GAF-F is limited. More study should be done of reliability and validity for both GAF-S and GAF-F scales individually.

FURTHER DEVELOPMENT FOR SCALE PROPERTIES

The history of GAF does not show the research-based development of GAF to be especially strong, particularly in the context of its widespread use. Little study of systematic variation in system properties has been carried out. Many alternative forms of a new GAF could be examined (with both with major and minor changes). It is difficult to forecast which changes are likely to provide the most significant improvements. For work with a new GAF, some overall goals can be formulated:

(1) GAF should continue to be an overall (i.e. global) measure of how patients are doing; (2) with a future GAF, it should be possible to rate severity from the most severe mental illness to perfect health; (3) GAF should continue not be a diagnosis-specific scoring system, but the generic properties should be improved; (4) results from GAF scoring should continue to add information compared to what diagnoses give; (5) for a new GAF, reliability should not be lower, but rather improved; (6) work with a new GAF should aim at improved validity; (7) sensitivity should be analysed, compared to other scaling methods, and found to be good enough for the purpose; (8) clinicians should find a new GAF to make sense; and (9) scoring with a new GAF should be little work requiring, i.e. scoring

should be fast and easy. The goals are ambitious, but not necessarily impossible to combine.

CONCLUSIONS

No doubt, GAF has a history with limited change of basic properties. It is too simple to believe that improvement work should not be done because GAF is good enough. An international research programme with study of effects of different changes in basic properties may well be important, but is lacking. Research on basic properties has not at all played an important role for further development of GAF. Problems with GAF may be related to this. Future research could improve GAF.

REFERENCES

Aas, I. H. M. (1991). Poliklinikker og dagkirurgi. Virksomhetsbeskrivelse for ambulant helsetjeneste [Outpatient clinics and daysurgery. Describing the activity of ambulatory care]. Goteborg: Nordic School of Public Health, NHV-rapport 1991, 4.

Aas, I. H. M. (2010). Global Assessment of Functioning (GAF): Properties and frontier of current knowledge. *Annals of General Psychiatry*, *9*, 20.

Aas, I. H. M. (2011). Guidelines for rating Global assessment of Functioning (GAF). *Annals of General Psychiatry*, *10*, 2.

Aas, I. H. M. (2014). Collecting Information for rating Global Assessment of Functioning (GAF): Sources of Information and Methods for Information Collection. *Current PsychiatryReviews*, *10*, [EPub Ahead of Print].

Bacon, S. F., Collins, M. J., & Plake, E. V. (2002). Does the Global Assessment of Functioning assess functioning? *Journal of Mental Health Counseling*, *24*, 202-212.

Bates, L. W., Lyons, J. A., & Shaw, J. B. (2002). Effects of brief training on application of the global assessment of functioning scale. *Psychological Reports*, *91*, 999-1006.

Bech, P., Malt, U. F., Dencker, S. J., Ahlfors, U. G., Elgen, K., Lewander, T., et al. (1993). Scales for assessment of diagnosis and severity of mental disorders. *Acta Psychiatrica Scandinavica*, *87*(Suppl 372), 3-86.

Bird, H. R., Canino, G., Rubio-Stipec, M., & Ribera, J. C. (1987). Further measures of the psychometric properties of the Children's Global Assessment Scale. *Archives of General Psychiatry*, *44*, 821-824.

Bowling, A. (1993). Measuring health. A review of quality of life measurements scales. Buckingham: Open University Press.

Bowling, A. (1997). Measuring disease. A review of disease-specific quality of life measurement scales. Buckingham: Open University Press.

Breakwell, G., & Millward, L. (1995). Basic evaluation methods. Leicester: British Psychological Society Books.

Bruyn, E. E. J. (2003). Assessment process. In: Fernández-Ballesteros R, editor. *Encyclopedia*

of Psychological Assessment. Thousand Oaks: Sage, pp. 93-7.

Burlingame, G. M., Dunn, T. W., Chen, S., Lehman, A., Axman, R., Earnshaw, D., et al. (2005). Selection of outcome assessment instruments for in patients with severe and persistent mental illness. *Psychiatric Services*, *56*, 44-451.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale; a procedure for measuring overall severity of psychiatric disturbance. Arch Gen Psychiatry, 33, 766-771.

Fallmyr, O., & Repal, A. (2002). Evaluering av GAF-skaring som del av Minste Basis Datasett [Evaluation of GAF-scoring as part of Minimum Basis Dataset]. *Tidsskr Nor Psykologforening*, *39*, 1118-1119.

Feinstein, A. R., Josephy, B. R., & Wells, C. K. (1986). Scientific and clinical problems in indexes of functional disability. *Annals of Internal Medicine*, *105*, 413-420.

First, M. B. (1995). Mastering DSM-IV Axis V. *Journal of Practical Psychiatry & Behavioural Health*, *1*, 258-259.

Gelder, M., Mayou, R., & Geddes, J. (2006). *Psychiatry, 3rd ed.* Oxford: Oxford University Press.

Goldman, H. H., Skodol, A. E., & Lave, T. R. (1992). Revising axis V for DSM-IV: a review of measures of social functioning. *The American Journal of Psychiatry*, *149*, 1148-1156.

Gonella, J. S. (1983). *Clinical criteria for disease staging*. Santa Barbara: Systemetrics Inc.

Hall, R. C. W. (1995). Global Assessment of functioning. A modified scale. *Psychosomatics*, *36*, 267-275.

Hay, P., Katsikitis, M., Begg, J., Da Costa, J., & Blumenfeld, N. (2003). A two-year follow-up study and prospective evaluation of the DSM-IV Axis V. *Psychiatry Services*, *54*, 1028-1030.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Bauman, B. D., Baity, M. R., Smith, S. R., et al. (2000). Reliability and validity of DSM-IV axis V. *The American Journal of Psychiatry*, *157*, 1858-1863.

Jones, S. H., Thorncroft, G., Coffey, M., & Dung, G. (1995). A brief mental health outcome scale reliability and validity of the Global Assessment of Functioning (GAF). *The British Journal of Psychiatry*, *166*, 654-659.

Karterud, S., Pedersen, G., Lovdal, H., & Friis, S. (1998). S-GAF. Global funksjonsskaring – splittet versjon (Global Assessment of Functioning – Split version). Bakgrunn og skaringsveiledning. Oslo: Klinikk for Psykiatri, Ulleval sykehus.

Lingjaerde, O., Bech, P., Malt, U., Dencker, S. J., Elgen, K., & Ahlfors, U. G. (1989). Skalaer for diagnostikk og sykdomsgradering ved psykiatriske tilstander. Del 1: Metodologiske aspekter [Diagnostic scales and disease grading in psychiatry. Part 1: Methodologic aspects]. *Nordic Journal of Psychiatry*, *43*(19), 1-39.

Loevdahl, H., & Friis, S. (1996). Routine evaluation of mental health: reliable information or worthless 'guesstimates'? *Acta Psychiatrica Scandinavica*, *93*, 125-128.

McDowell, I., & Newell, C. (1987). *Measuring health: a guide to rating scales and questionnaires*. Oxford: Oxford University Press.

Moos, R. H., McCoy, L., & Moos, B. S. (2000). Global Assessment of Functioning (GAF) ratings: determinants and role as predictors of one-year treatment outcomes. *Journal of Clinical Psychology*, *56,* 449-461.

Moos, R. H., Nichol, A. C., & Moos, B. S. (2002). Global Assessment of Functioning ratings and the allocation and outcomes of mental health services. *Psychiatry Services*, *53*, 730-737.

Niv, N., Cohen, A. N., Sullivan, G., & Young, A. (2007). The MIRECC Version of the Global assessment of Functioning scale: Reliability and validity. *Psychiatry Services*, *58*, 529-535.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill Inc.

Parker, G., O'Donell, M., Hadzi-Pavlovic, D., & Roberts, M. (2002). Assessing outcome in community mental health patients: a comparative analysis of measures. *International Journal of Social Psychiatry*, *48*(1), 11-19.

Patterson, D. A., & Lee, M. S. (1995). Field trial of the Global Assessment of Functioning Scale – Modified. *The American Journal of Psychiatry*, *152*, 1386-1388.

Pedersen, G., Hagtvedt, K. A., & Karterud, S. (2007). Generalizability studies of the Global Assessment of Functioning – split version. *Comprehensive Psychiatry*, *48*, 88-94.

Piersma, H. L., & Boes, J. L. (1997). The GAF and psychiatric outcome: a descriptive report. *Community Mental Health Journal*, *33*, 35-41.

Rey, J. M., Starling, J., Weaver, C., Dossetor, D. R., & Plapp, J. M. (1995). Inter-rater reliability of global assessment of functioning in a clinical setting. *Journal of Child Psychology Psychiatry*, *36*, 787- 792.

Robert, P., Aubin, V., Dumarcet, M., Braccini, T., Souetre, E., & Darcourt, G. (1991). Effect of symptoms on the assessment of social functioning: comparison between Axis V of DSM III-R and the psychosocial aptitude rating scale. *European Psychiatry*, *6*, 67-71.

Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York: The Guilford Press.

Rosse, R. B., & Deutsch, S. I. (2000). Use of the Global Assessment of Functioning scale in the VHA: Moving toward improved precision. *Veterans Health System Journal*, *5*(4), 50-58.

Roy-Byrne, P., Dagadakis, C., Unutzer, J., & Ries, R. (1996). Evidence for limited validity of the revised Global Assessment of Functioning Scale. *Psychiatry Services*, *47*, 864-866.

Salvi, G., Leese, M., & Slade, M. (2005). Routine use of mental health outcome assessments: choosing the measure. *The British Journal of Psychiatry*, *186*, 146-152.

Schorre, B. E. H., & Vandvik, I. H. (2004). Global assessment of psychosocial functioning in child and adolescent psychiatry. A review of three unidimensional scales (CGAS, GAF, GAPD). *European Child & Adolescent Psychiatry*, *13*, 273- 286.

Schrader, G., Gordon, M., & Harcourt, R. (1986). The usefulness of DSM-III Axis IV and Axis V assessments. *The American Journal of Psychiatry*, *143*, 904-907.

Sederer, L. I., Herman, R., & Dickey, B. (1995). The imperative of outcome assessment in psychiatry. *American Journal of Medical Quality*, *10*, 127-132.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology. An Introduction. *American Psychologist*, *55*, 5-14.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress. Empirical validation of interventions. *American Psychologist*, *60*, 410-421.

Soderberg, P., Tungstrom, S., & Armelius, B. A. (2005). Reliability of Global Assessment of Functioning ratings made by clinical psychiatric staff. *Psychiatry Services*, *56*, 434-438.

Startup, M., Jackson, M.C., & Bendix, S. (2002). The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology*, *41*, 417-422.

Streiner, D. L., & Norman, G. R. (19094). Health Measurement scales. A practical guide to their development and use. Oxford: Oxford University Press.

Sutherland, H. J., Dunn, V., & Boyd, N. F. (1983). Measurement of values for states of health with linear analog scales. *Medical Decision Making*, *3*, 477-87.

Thomson, C. (1989). Introduction. In The instruments of psychiatric research. Edited by Thomson C. Chichester: John Wiley & Sons, pp 1-17.

Tungstrom, S., Soderberg, P., & Armelius, B. A. (2005). Relationship

between the Global Assessment of Functioning and other DSM Axes in routine clinical work. *Psychiatry Services*, 56, 439-443.

Vaillant, G. E. (2003). Mental health. *The American Journal of Psychiatry*, 160, 1373-1384.

Vatnaland, T., Vatnaland, J., Friis, S., & Opjordsmoen, S. (2007). Are GAF scores reliable in routine clinical use? *Acta Psychiatrica Scandinavica*, 115, 326-330.

Wells, K. B., Stewart, A., Hays, R. D., Burnam, M. A., Rogers, W., Daniels, M., et al. (1989). The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *Journal of American Medical Association*, 262, 914-919.

Widiger, T. A., & Clark L. E. (2000). Toward DSM-V and the classification of psychopathology. *Psychological Bull*, 126, 946-963.

Young, F. W. (1984). Scaling. *Annual Reviews of Psychology*, 35, 55-81.