



ELSEVIER

Child Abuse & Neglect 31 (2007) 909–915

Child Abuse  
& Neglect

Brief communication

## Reliability of the GAF and CGAS with children exposed to trauma<sup>☆</sup>

Katherine Blake<sup>a,\*</sup>, Scott Cangelosi<sup>a</sup>,  
Sonja Johnson-Brooks<sup>a</sup>, Harolyn M.E. Belcher<sup>a,b,c,d</sup>

<sup>a</sup> *The Kennedy Krieger Institute Family Center, 2901 E. Biddle Street,  
Baltimore, MD 21213, USA*

<sup>b</sup> *The Kennedy Krieger Institute, Department of Neurology and Child Development,  
Baltimore, MD, USA*

<sup>c</sup> *The Johns Hopkins School of Medicine, Department of Pediatrics,  
Baltimore, MD, USA*

<sup>d</sup> *The Johns Hopkins Bloomberg School of Public Health, Department of Mental Health,  
Baltimore, MD, USA*

Received 2 October 2005; received in revised form 22 October 2006; accepted 21 December 2006

---

*Keywords:* GAF; CGAS; Children; Complex trauma

---

### Introduction

The latest data from the National Child Abuse and Neglect Data System indicates that an estimated 872,000 children in the United States were victims of abuse or neglect in the year 2004 (US Department of Health and Human Services, 2006). The physical and mental health consequences of recurrent, often parent-inflicted, child maltreatment are substantial (DosReis, Zito, Safer, & Soeken, 2001; Dube et al., 2003; Felitti et al., 1998). Sequelae of maltreatment have been described as “an environmentally induced complex developmental disorder” (De Bellis, 2001) that may extend beyond posttraumatic stress disorder

---

<sup>☆</sup> This study was supported in part by a grant from the Substance Abuse and Mental Health Services Administration, 5 U 79 SM56215 and the National Institute on Mental Health, MH01882.

\* Corresponding author.

(PTSD) symptoms to include multiple domains of impairment including attachment, biology, affect regulation, dissociation, behavior regulation, cognition, and self-concept (DosReis et al., 2001; Dube et al., 2003; Felitti et al., 1998).

The two most commonly used mental health measures of functioning of traumatized and non-traumatized children are the Global Assessment of Functioning (GAF) scale and the Children's Global Assessment Scale (CGAS, Shaffer et al., 1983). The GAF was first introduced as Axis V in the Diagnostic and Statistical Manual of Mental Disorders Third Edition Revised (DSM-III-R; American Psychiatric Association, 1987). The CGAS was developed later as a more child-specific measure of functioning. These measures are currently used in diagnosis, treatment, and evaluation of children's mental health problems to determine eligibility to receive mental health services and document treatment outcome status (Bates, 2001).

Only a limited number of studies have examined the reliability of the CGAS and GAF with child populations. Studies of the reliability of these measures have yielded mixed results. Estimates of interrater reliability of the CGAS have ranged from .53 to .93 (Dyrborg et al., 2000; Green, Shirk, Hanze, & Wanstrath, 1994; Rey, Starling, Wever, Dossetor, & Plapp, 1995; Shaffer et al., 1983; Steinhausen, 1987). Likewise, studies of the interrater reliability of the GAF for children and adolescents yielded coefficients ranging from .54 to .92 (Beitchman et al., 2001; Gold, Sherra, & Clarkson, 1993; Manassis & Hood, 1998; Rey et al., 1995; Smith, Thienemann, & Steiner, 1992).

A recent literature review yielded no studies that specifically examined the interrater reliability of the GAF and CGAS for the evaluation of children with trauma. The complex presentation of children exposed to maltreatment may increase the likelihood that traditional measures may be less reliable. It is particularly important to reliably assess the functioning of this population given the substantial negative consequences of trauma (Anda et al., 2006). The aim of the present study was to estimate the reliability of the GAF and the CGAS using clinical case vignettes of both traumatized and non-traumatized children. It was unclear whether the GAF and CGAS would yield reliability estimates with a population of traumatized children that were similar to those obtained in previous studies with child mental health treatment populations without a history of trauma. Since the CGAS has descriptions that more specifically reflect child functioning, it was hypothesized that clinicians would be more reliable when rating case vignettes using the CGAS as compared to the GAF, particularly with vignettes of traumatized children.

## Method

### *Study sample selection*

Participants for this study were recruited from the staff of an urban community-based mental health center (Center) specializing in the treatment of children with complex trauma. All raters were practicing clinicians familiar with the use of the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition Text Revision (DSM-IV-TR; American Psychiatric Association, 2000) and level-of-functioning measures, including the GAF. Clinicians volunteered to participate in the study. Clinicians participating in the study signed an informed consent approved by the Johns Hopkins Medicine Institutional Review Boards.

### *Measures*

The GAF is a scale of 1–100 divided into 10-point increments to delineate increasingly higher functioning as the rating approaches 100. “Anchor points” separate increments into “superior functioning” (91–100), “good functioning” (81–90), “no more than slight impairment” (71–80), “mild symptoms” (61–70), “moderate symptoms” (51–60), “serious symptoms” (41–50), and so forth. As a diagnostic axis, the GAF has descriptors that apply to both children and adults.

The C-GAS is a comparable scale of 1–100 with similarly divided increments of functioning. The C-GAS descriptors are specifically targeted to children 4 through 16 years of age. The psychometric properties of the CGAS and GAF have been previously described in this paper.

### *Clinical vignettes*

Clinical vignettes were selected from three different sources to include a mix of ages (4–15 years), gender, diagnostic formulations, and case presentation formats. Ten of the vignettes were based on clinic intake interviews performed at the Center. Five of the children in the Center clinic vignettes had a history of complex trauma exposure (Center-T vignettes). The cases were as follows: 5-year-old male with history of physical abuse and neglect, 6-year-old female with history of neglect, sexual abuse, and exposure to domestic violence, 9-year-old female with history of physical abuse and neglect, 10-year-old male with history of physical abuse and exposure to domestic violence, and 14-year-old female with history of physical and sexual abuse. The other five Center intake interviews had similar types of pathology, but did not include a history of complex trauma in the clinical presentation (Center-NT vignettes). The final five vignettes were selected from the DSM-IV-TR Casebook (Spitzer, Gibbon, Skodol, Williams, & First, 2002) and the DSM-IV-TR Case Studies (Frances & Ross, 2001). These vignettes were specifically chosen for the absence of trauma (DSM-NT vignettes).

### *Procedure*

Clinicians rated the 15 clinical case vignettes using both the GAF and the CGAS. Clinicians were instructed to review the scoring rules for both the GAF and the CGAS scales prior to reading the vignettes. They were then given written vignettes and asked to determine the patient’s level of functioning described in the clinical vignette. Clinicians were randomly assigned to use either the CGAS or the GAF, and the vignette presentation order was also randomized. Rating instruments were alternated such that, clinicians who initially rated the vignettes using the CGAS, used the GAF 1 month later to rate the same vignettes, and clinicians who were initially randomized to use the GAF to rate vignettes used the CGAS 1 month later.

### *Data analysis*

Descriptive information regarding the participants was collected including type of graduate degree, years of postgraduate clinical experience, and years of postgraduate clinical experience working with children with trauma. Interrater reliability was calculated using intraclass correlation (ICC) to measure the reproducibility of replicate measures from the same subject. ICC ranges from 0 and 1.0, with 0 indicating no reproducibility and 1.0 indicating perfect reproducibility (Rosner, 2000). An ICC value less

than .4 suggest poor reproducibility, .4 to less than .75 indicates fair to good reproducibility, and .75 or greater is consistent with excellent reproducibility (Fleiss, 1986). The model assumed that rater effects are random and measures effects are fixed (two-way mixed effects model); confidence intervals for the ICC are presented based on the methods given by Shrout and Fleiss (1979). Fisher's Z-transformations were conducted and pairwise comparisons of ICCs for DSM-NT vignettes and Center-NT vignettes to Center-T vignettes were performed for both the GAF and CGAS. To maintain the overall significance level ( $\alpha = .05$ ) for comparing the three vignettes a Bonferroni correction was used (adjusted  $\alpha = .0125$ ). In addition, data from the DSM-NT vignettes and Center-NT vignettes were pooled and compared to the Center-T vignettes for both the GAF and CGAS (adjusted  $\alpha = .025$ ). Univariate analyses of variances were also conducted in order to examine the impact of level of clinician experience and order of assignment of measures on ratings.

## Results

### *Clinician raters*

There were 22 clinician raters who included licensed clinical social workers ( $n = 17$ ), doctoral level psychologists ( $n = 2$ ), and master's level psychology associates ( $n = 3$ ). The clinicians' experience ranged from 6 months to 22 years.

### *Interrater reliability*

Table 1 displays the interrater reliability scores with 95% confidence intervals. Clinicians had good interrater reliability when rating vignettes of children without trauma histories on both the GAF and

Table 1  
Interrater reliability

Vignette type	DSM-IV-TR casebook without trauma history (DSM-NT) ( $N = 5$ )	Clinic intake without trauma history (Center-NT) ( $N = 5$ )	All cases without trauma history ( $N = 10$ )	Clinic intake with trauma history (Center-T) ( $N = 5$ )
# of clinicians				
GAF	22	18	18	22
CGAS	20	19	18	21
Mean score ( <i>SD</i> )				
GAF	52.01 (5.32)	55.08 (4.25)	53.46 (4.31)	38.62 (8.25)
CGAS	51.50 (7.20)	56.43 (5.12)	53.95 (5.17)	41.49 (7.16)
ICC				
GAF	.69	.73	.68	.33
CGAS	.55	.60	.57	.38
95% C.I.				
GAF	.42–.95	.47–.96	.48–.88	.13–.81
CGAS	.28–.91	.33–.93	.36–.82	.16–.84

CGAS (.55–.73). The obtained ICCs were similar to those previously reported in the literature (e.g., Rey et al., 1995). In contrast, clinicians displayed poor reliability using the GAF and CGAS in rating functional status of children with a history of trauma (.33–.38).

Estimates of GAF and CGAS reliability for children with and without histories of trauma were compared using Fisher's Z-transformation for ICC. Center-T vignettes were compared to Center-NT vignettes and DSM-NT vignettes in a pairwise fashion. There were no statistically significant differences between the reliability for either the GAF or CGAS when comparing the Center-T vignettes to Center-NT vignettes ( $Z = 1.20, p = .23$  for GAF;  $Z = .64, p = .52$  for CGAS). There was also no significant difference between the reliability of the GAF or CGAS when comparing Center-T to the DSM-NT vignettes ( $Z = .22, p = .82$  for GAF;  $Z = .38, p = .70$  for CGAS). The Center-NT and DSM-NT vignettes were combined to create a pooled non-trauma group. No statistically significant differences were found between the reliability for either the GAF or CGAS, comparing the pooled non-trauma group with the Center-T ( $Z = 1.18, p = .24$  for GAF;  $Z = .64, p = .52$  for CGAS). Vignette source (Center-NT vs. DSM-NT) also did not significantly impact the results on either the GAF ( $Z = .02, p = .98$ ) or the CGAS ( $Z = .12, p = .90$ ).

Univariate analyses of variances were conducted in order to determine whether there was any impact of level of clinician experience on the results. Number of years of clinical experience did not significantly impact either GAF or CGAS scores. There was also no significant effect based upon which measure the clinician was assigned (GAF or CGAS) during the first rating of a vignette.

## Discussion

Raters display poor reliability in rating the functional status of children with a history of trauma. In contrast, clinicians display good interrater reliability when rating vignettes of children without trauma histories on both the GAF and CGAS. Contrary to our hypothesis, the CGAS was not more reliable than the GAF on either the trauma or non-trauma vignettes. On non-trauma vignettes, reliability estimates obtained were consistent with previous research on the reliability of use of the GAF and CGAS with children (e.g., Rey et al., 1995).

Reliable rating of the functional status of children in mental health treatment is of utmost importance for monitoring children's clinical progress and demonstrating treatment effectiveness. The GAF is the most commonly used single score instrument designed to measure functional status of patients engaged in mental health treatment. The CGAS was developed as a single score measure of children's functioning. Reliability of these measures for children with a history of complex trauma exposure has not been previously published.

Interrater reliability was lower on clinical vignettes describing children with a history of trauma when compared to vignettes without a history of child maltreatment using both the CGAS and GAF. While this finding did not reach statistical significance, it is important to note that, regardless of the source of clinical vignettes (actual clinic intakes or DSM-IV-TR case studies), the pattern of findings remained unchanged. Vignettes with trauma demonstrated less reliability among the clinical raters. The findings suggest that the unique, dynamic characteristics inherent in trauma case descriptions, results in lower reliability.

The difference in interrater reliability between the vignettes with and without trauma did not reach statistical significance. The lack of statistical significance between the reliability scores may be attributed to the variability present on both the GAF and CGAS scores for cases with trauma. The larger variability seen in trauma cases may be due to the dynamic and complex presentation of children with a history of

trauma. The results suggest that GAF and CGAS may not reliably measure functioning in children with complex trauma exposure.

Limitations of the study include the number of clinical trauma vignettes used in this preliminary research. Replication of the study using more trauma case vignettes and multiple sites is needed to confirm whether the reliability and variance demonstrated in this study is supported when larger numbers of raters and cases are used. Another limitation of the study is that clinicians used written case vignettes. The true application of any level-of-functioning measure occurs with live clients rather than written vignettes. Replication of this study with actual clients using videotaped assessments may further validate findings of interrater reliability of child trauma cases.

Strengths of this study include the use both of standardized vignettes and actual clinical intakes from children with and without trauma. In addition, the sample of clinician raters from a center specializing in the assessment and treatment of children with maltreatment is important in determining the usefulness of the CGAS and GAF for children exposed to trauma. If variability in scoring exists in a specialized mental health setting, less experienced clinicians may also demonstrate less reliability on cases involving trauma.

Clinicians often use the GAF and CGAS to help make decisions about treatment progress, outcomes, and needs. Results of this study suggest variability in clinicians' reliability, favoring child cases without a history of trauma compared to those with a trauma history. Clinicians should exercise caution when using these measures to evaluate the functional status of children exposed to complex trauma. These measures may not adequately characterize the functional status of traumatized children; possibly leading to an over or under-representation of the child's functional status.

Interrater reliability of the two most popular level-of-functioning measures, the GAF and the CGAS, may be affected by the clinical presentation of trauma. A single scale of 1–100 may inadequately capture the complex symptoms that are present when children have experienced chronic trauma throughout early childhood. Our results suggest the need to study these findings on a larger scale.

## References

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., Rev.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Rev.). Washington, DC: Author.
- Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, C., Perry, B. P., Dube, S. R., & Giles, W. H. (2006). The enduring effects of abuse and related experiences in childhood. *European Archives of Psychiatry and Clinical Neuroscience*, 256, 174–186.
- Bates, M. P. (2001). The child and adolescent functional assessment scale (CAFAS): Review and current status. *Clinical Child and Family Psychology Review*, 4, 63–84.
- Beitchman, J. H., Adlaf, E. M., Douglas, L., Atkinson, L., Young, A., Johnson, C. J., Escobar, M., & Wilson, B. (2001). Comorbidity of psychiatric and substance use Disorders in late adolescence: A cluster analytic approach. *American Journal of Drug and Alcohol Abuse*, 27, 421–440.
- De Bellis, M. D. (2001). Developmental traumatology: The psychobiological development of maltreated children and its implications for research, treatment, and policy. *Development and Psychopathology*, 13, 539–564.
- DosReis, S., Zito, J., Safer, D., & Soeken, K. L. (2001). Mental health services for youths in foster care and disabled youths. *American Journal of Public Health*, 91(7), 1094–1099.
- Dube, S. R., Felitti, V. J., Dong, M., Chapman, D. P., Giles, W. H., & Anda, R. F. (2003). Childhood abuse, neglect, and household dysfunction and the risk of illicit drug use: The adverse childhood experience study. *Pediatrics*, 111(3), 564–572.

- Dyrborg, J., Larsen, F. W., Nielsen, S., Byman, J., Nielsen, B. B., & Gautre-Delay, F. (2000). The children's global assessment scale (CGAS) and global assessment of psychosocial disability (GAPD) in clinical practice—substance and reliability as judgment by intraclass correlations. *European Child and Adolescent Psychiatry*, 9, 195–201.
- Felitti, V., Anda, R., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences study (ACE). *American Journal of Preventative Medicine*, 14(4), 245–258.
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Frances, A., & Ross, R. (2001). *DSM-IV-TR case studies A clinical guide to differential diagnosis*. Washington, DC: American Psychiatric Publishing.
- Gold, J., Shera, D., & Clarkson, B., Jr. (1993). Private psychiatric hospitalization of children: Predictors of length of stay. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 135–143.
- Green, B., Shirk, S., Hanze, D., & Wanstrath, J. (1994). The Children's Global Assessment Scale in clinical practice: An empirical evaluation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 1158–1164.
- Manassis, K., & Hood, J. (1998). Individual and familial predictors of impairment in childhood anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 428–434.
- Rey, J. M., Starling, J., Wever, C., Dossetor, D. R., & Plapp, J. M. (1995). Inter-rater reliability of global assessment of functioning in a clinical setting. *Journal of Child Psychology and Psychiatry*, 36, 787–792.
- Rosner, B. (2000). *Fundamentals of biostatistics* (5th ed.). Pacific Grove: Duxbury Press.
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, 40, 1228–1231.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin*, 86, 420–428.
- Smith, C., Thienemann, M., & Steiner, H. (1992). Defensive style and adaptation in adolescents with depressive and eating disorders. *Acta Paedopsychiatrica*, 55, 185–186.
- Spitzer, R. L., Gibbon, M., Skodol, A. E., Williams, J. B. W., & First, M. B. (2002). *DSM-IV-TR case book. A learning companion to the Diagnostic and Statistical Manual of Mental Disorders* (4th ed., Text rev.). Washington, DC: American Psychiatric Publishing.
- Steinhausen, H. C. (1987). Global assessment of child psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 203–206.
- US Department of Health and Human Services, Administration on Children, Youth, and Families (2006). *Child Maltreatment 2004*. Washington, DC: US Government Printing Office.