Published in final edited form as:

J Nerv Ment Dis. 2014 November ; 202(11): 822–828. doi:10.1097/NMD.0000000000000200.

Assessing overall functioning with adolescent inpatients

Greg Haggerty, Ph.D.¹, Nicholas Forlenza, Ph.D.¹, Charlotte Poland, MSW¹, Sagarika Ray, MD¹, Jennifer Zodan, Ph.D.¹, Ashwin Mehra, Ph.D.¹, Ajay Goyal, MD¹, Matthew R. Baity, Ph.D.², Caleb J. Siefert, Ph.D.³, Sean Sobin, M.A.⁴, David Leite, B.S.⁵, and Samuel J. Sinclair, Ph.D.⁶

¹Von Tauber Institute for Global Psychiatry, Nassau University Medical Center, East Meadow, NY, USA

²Alliant International University, Sacramento, CA, USA

³University of Michigan-Dearborn, Dearborn, MI, USA

⁴Long Island University-CW Post Campus, Brookville, NY, USA

⁵Liberty University, Lynchburg, VA, USA

⁶Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Abstract

The current study looks to evaluate the validity and reliability of a brief measure of overall functioning for adolescents. Clinicians were asked to complete the Overall Functioning Scale for 72 adolescents consecutively admitted to the adolescent psychiatric inpatient service of a community safety net medical center. The results revealed that this new measure is related to the patients' length of stay, clinician-rated measures of social cognition and object relations, Global Assessment of Functioning score at admission, and global rating of engagement in individual psychotherapy. Results also showed that the OFS was related to patients' history of non-suicidal self-harm as well as treatment outcome as assessed by measures of psychological health and well-being and symptomatology. Hierarchical regressions reveal that the OFS shows incremental validity above the admission GAF score in predicting length of stay. The results also showed that the OFS demonstrates inter-rater reliability in the excellent range (ICC_{1,2}) of .88. Clinical implications of the use of this tool as well as areas of future research are discussed.

Keywords

Overall functioning; GAF; SOS-10; SCORS-G; adolescent; inpatient; Assessment

The monitoring of outcomes for child and adolescent mental health services has become an increasing focus both scientifically and economically (Kazdin, 2003; Williams & Kerfoot, 2005; Epstein, Kutash, & Duchnowski, 2005). Regulatory agencies and insurance providers recommend that measuring outcome be part of routine clinical practice (Weiss, 1998). This

is even more important for adolescent inpatient treatment which is often the most costly and disruptive to the child's life. One important aspect of assessing outcome for adolescents and adults is measuring their overall functioning, aside from symptomatology. This functional rating should take into consideration a variety of important domains. An adolescent's ability to function well at school, show appropriate and adaptive peer and family relations is important to the assessment of outcome of any treatment (Winters, Collett, & Myers, 2005). This paper looks to investigate a brief, unifactoral measure of global functioning to be used with adolescent.

Existing Unidimensional Functioning Measures for Adolescents

Changes made in the DSM have resulted in the elimination of the multiaxial diagnosis which included the Global Assessment of Functioning (GAF; American Psychiatric Association, 1994). As a result, many clinicians in regular practice are not assessing functioning in a quantifiable way. Clinicians need a measure that can be scored validly and reliably, is easy to interpret/use and conveys information about the adolescent's functioning over several core domains.

Aside from the GAF, clinicians and researchers have also used the Children's Global Assessment Scale (CGAS; Shaffer et al., 1983) which is an adaptation from the GAF. Both the CGAS and GAF are scales ranging from 1 (lowest functioning) to 100 (excellent functioning) with 10 anchors representing deciles (e.g., 1–10, 11–20, 91–100). Each 10-point decile is anchored by a brief paragraph describing characteristics of functioning at that level and, like the GAF, raters are instructed to rate the lowest level of overall functioning for that adolescent during the specific time period. One limitation with both the GAF and CGAS is that they provide one number for all aspects of functioning. The reader would not be able to discern if a patient was functioning poorer in school than in their interpersonal relationships (Winters et al., 2005).

Other unifactoral measures of functioning for adolescents include the Columbia Impairment Scale (CIS; Bird, 1999) which is a 13-item respondent-based measure and the Brief Impairment Scale (BIS; Bird, Canino, Davies, Ramirez, Chavez, Duarte, & Shen, 2005), a 23-item respondent-based measure which builds on the CGAS and CIS as it is a brief. However, the BIS is slightly longer than the CIS and attempts to assess multiple domains of functioning independently of symptomatology. Both the CIS and BIS may not be as appropriate for inpatient use. A measure suitable for inpatient assessment would have to be quick to use, easy to understand, provide a good amount of information about the adolescent's functioning on several core domains, valid and reliable. The measure would also be useful if it were able to relate to important hospital and unit metrics such as length of stay.

New Measure of Overall Functioning

The new measure of overall functioning investigated in this study looks to assess adolescents globally using three domains of adolescent functioning: chronic or general level of personality functioning, quality of peer relationships and school performance. These three domains of functioning were included in this measure from a review of the literature on the

assessment of functioning for adolescents (Winters et al., 2005). These three constructs have been used in other global functioning measures such as the CGAS and the GAF (Bird, 1999; Bird, Canino, Rubio-Stipec, & Ribera, 1987; Winters, Collett, & Myers, 2005).

Additionally, the inclusion of a dimensional assessment of personality functioning into the OFS was supported by work showing how personality pathology affects overall functioning (Kuermann & Hilsenroth, 2012; Peters, Hilsenroth, Eudell-Simmons, Blagys, & Handler, 2005; Stein, Hilsenroth, Pinsker-Aspen, & Primavera, 2009; Westen & Arkowitz-Westen, 1998; Zittel-Conklin, & Westen, 2005). Personality pathology in adolescents has also been linked with other psychiatric disorders (for review Shiner & Tackett, 2013).

In the current study, we investigated the construct validity and reliability of a new measure designed to assess global functioning for inpatient adolescents, the Overall Functioning Scale. We anticipate the OFS will be positively correlated with length of stay, clinician-ratings on a measure of object relations: Social Cognition and Object Relations Scale-Global Scoring Version (SCORS-G; Stein, Hilsenroth, Slavin-Mulford, & Pinsker, 2011; Westen, 1995), Global Assessment of Functioning scores, clinician-ratings of overall engagement in individual and group therapy, history of nonsuicidal self-injury and eating disorder behavior, and changes in outcome measures from admission to discharge. We did not have any inter-rater reliability data for the initial study sample because the OFS ratings were made as part of the patients' medical chart. As a result, we conducted a second study using a different sample of adolescent inpatients from the same setting which solely investigated the inter-reliability of the OFS.

Previous research has shown that assessments of overall functioning are related to patient's length of stay on an inpatient service. This is especially important as administrators, payers and health service researchers are increasingly interested in length of stay as an indicator of the quality and efficiency of inpatient psychiatric hospitalization (Blais et al., 2003; Compton, Craw, & Rudisch, 2006). Both studies investigated variables related to length of stay including the GAF. Lower scores on the GAF were found to be related to longer length of stays for patients. Gold, Shera, and Clarkson (1992) studied predictors of length of stay of a children's psychiatry inpatient service and found that overall functioning as assessed by the CGAS was a "powerful predictor" of length of stay. Gordon et al. (1985) also found this result in adolescents and adults as well.

Defife, Goldberg, & Westen (2013) found that global functioning was related to ratings of patients' object relations in a sample of adolescents being seen for outpatient treatment. Specifically, the results showed that adolescents who demonstrated lower ratings of self and interpersonal functioning as measured by the Social Cognition and Object Relations Scale-Global Rating Method (SCORS-G; Stein, Hilsenroth, Slavin-Mulford, & Pinsker, 2011) also showed lower ratings of overall functioning which include school functioning, peer relations and general personality functioning.

Roedelof, Ilja, & Bongers (2013) have found that treatment engagement has been related to patients' overall functioning in a sample of adolescents with severe psychiatric problems. In this study, they separated their sample of patients into those low, medium and high on

engagement in treatment. Their results showed that low engagement group had an average GAF score that was lower than the other two groups.

Nonsuicidal self-harm (NSSI) has also been linked with overall functioning. Selby, Bender, Gordon, Nock and Joiner (2012) looked at the relationship between NSSI and patients with and without the diagnosis of Borderline Personality Disorder. The results showed that NSSI is linked with lower overall functioning as measured by the patients' GAF scores. Research (Wentz, Gillberg, Anckarsater, Gollberg, & Rastam, 2009) also shows that those who are diagnosed with eating disorders have lower GAF scores than those who did not.

STUDY 1: CONTRUCT VALIDITY

Methods

The sample consisted of 72 patients (45% of the total patients admitted to the unit), 52.8% male, consecutively admitted to the Adolescent Psychiatric Inpatient Unit of a large northeastern hospital. This study was approved by the hospital's IRB of record. Patients were between the ages of 13–17 with a mean age of 15.7 ($SD = 1.18$). Ethnic makeup of the sample was as follows: 40.8% Caucasian, 25.4% African American, 25.4% Latino/Hispanic/Spanish, 5.6% Other and 2.8% Asian. The primary diagnoses for these 72 patients were as follows: 64% Mood Disorders, 30% Conduct Disorder/Oppositional Defiant Disorder, 2% Impulse Control Disorder, 3% Psychosis, 1% Post Traumatic Stress Disorder. We found no significant differences in age, gender, or diagnosis between those who consented to the study versus those who did not consent. Patients who showed cognitive impairment or IQ below 70 were excluded from this study (4 patients).

Procedure—Parents and legal guardians were approached for participation by the research team member who was not one of the clinicians in the study at the end of the initial family meeting. Consented patients completed a battery of self-reports within one day of admission to the adolescent inpatient unit which included the Schwartz Outcome Scale-10 (SOS-10; Blais et al., 1999) and the Brief Symptom Measure-25 (BSM-25; Blais et al., 2013). Individual and group psychotherapists, blind to each others' ratings and self-report data, completed a GAF assessment based on the admission presentation of the patient (this was based on the description of the patient from the ER records and physical assessment, the first psychiatry attending assessment note and the first nursing assessment note in the chart). At discharge, the patient's individual and group therapist, blind to each others' ratings and patient self-report questionnaire data, completed a battery of clinical ratings which included the SCORS-G. The individual therapist complete an engagement rating for individual therapy and the group therapist did the same for group therapy.

Setting—The adolescent inpatient unit in this study is a 12-bed locked facility at a large northeastern safety-net hospital. The unit provides treatment for adolescents in acute distress. The average length of stay for patients on this unit is 10.81 days ($SD = 5.23$). While on the unit, adolescents receive psychopharmalogical treatment as deemed necessary, 2–3 individual psychotherapy sessions per week, 3 general group sessions per week, a weekly structured anger management group, a weekly substance use psycho-education group, daily

academic programming, and daily recreation therapy. On average, participants received 3 individual therapy sessions ($SD = 1.56$) and 6 group session ($SD = 3.08$) during their stay.

All patients, whether they were enrolled in this study or not, received individual psychotherapy and group therapy as part of their treatment on the unit. Patients on the unit were assigned to an individual in an ecologically valid manner based on clinician availability and caseload. Although patients could have a few different therapists leading or co-leading groups on the unit, the group therapist assigned to complete assessments for the patient had to be a co-leader of at least 3 groups each week on the unit.

Study Therapists/Raters—Participating clinicians included one licensed clinical psychologist with over 5 years of experience post graduate and advanced training in assessment and 5 (4 females) advanced clinical psychology doctoral students (3 interns and 2 externs) who had completed advanced coursework in assessment at an APA-accredited clinical psychology PhD program and were supervised by a licensed clinical psychologist. The trainees' ages ranged from 27–35 years old with an average age of 29.3. Two of the three psychology interns had extensive training in the therapeutic model of assessment (TMA; Finn & Tonsager, 1997) and the study measures before their participation in this study. The psychology interns had 4.3 years of training and both psychology externs had 3 years of training before starting their rotation on the unit. All clinicians kept a caseload of 2 patients throughout their rotation on the unit. This caseload was not limited to consented patients. At any given time, the unit contained a licensed psychologist, a psychology intern and an extern. All study clinicians met on a bi-weekly basis for reliability training for clinical rating measure.

Reliability Training—All clinicians providing ratings for this study met every two weeks for reliability training as a group. These meetings were run by the first author. The training included review of training materials for all of the study measures at the beginning of the study. At the biweekly meetings, members would blindly assess a patient they all knew well who was not part of the study and scoring discrepancies would be discussed. The group would contact an expert on the measure to discuss the issues further to ensure accurate scoring of study measures if scoring questions arose. Training on the clinical ratings occurred throughout the study duration.

The patient's individual therapist utilized all available information (i.e., chart, direct interaction with patient, feedback from unit staff) to complete dimensional ratings of the patient's chronic level of personality, peer relationships, and school functioning. The therapists in this study were part of the unit clinical staff and, although added to the research protocol, were not aware of any of the study hypotheses. A member of the research staff who was not part of the clinical team reviewed the patient's chart and noted any history of eating disordered behavior and nonsuicidal self-injury.

SCORS-G and GAF ratings were completed by both the patients' individual and group therapist to provide inter-rater reliabilities for these measures. We averaged together the two clinical raters' scores to provide a composite score. Averaging raters scores together to create composite scores are more reliable than ratings of individual raters because individual

idiosyncrasies tend to cancel out and the average score of the clinical raters reflects the consensual wisdom (the principle of aggregation; see. e.g., Horowitz, Inouye, & Siegelman, 1979; Rushton, Brainerd, & Presley, 1983; Shedler, Karliner, & Katz 2003).

Measures

Overall Functioning Scale (OFS): The OFS was created from three ratings of global functioning on the Clinician Data Form-Adolescent Version (CDF-A; Westen, Shedler, Durrett, Glass, & Martens, 2003). The CDF-A is used to gather a wide range of demographic and diagnostic information about the patient. For this study we investigated the clinical validity of three of the items which comprise overall functioning in adolescents which include: Chronic level of personality functioning (1 = severe personality disorder to 5 = high functioning); quality of peer relationships (1 = very poor or absent to 5 = very good); and school functioning (1 = failing/dropped out to 5 = working to full potential). These ratings were completed by the patient's individual therapist based on all available information. The OFS was completed after the initial family meeting and psychiatric interview with the patient. All of which happen within the first two business days after admission.

Social Cognition and Object Relations Scale – Global Ratings (SCORS-G; Stein, Hilsenroth, Slavin-Mulford, & Pinsker, 2011): The SCORS-G is a clinician-rated measure of a patient's representations of self and significant others. The scale consists of eight theoretically constructed variables that are scored on a 7-point Likert-type scale, where a lower score indicates more pathological responses and a higher score suggests healthy functioning. The eight variables are as follows: 1) Complexity of Representation (COM) reflects the richness of one's representations of self and others, one's ability to recognize internal states in self and others, and one's ability to integrate both positive and negative aspects of self and others; 2) Affective Quality of Representations (AFF) assesses one's expectations of others (positive or negative) and one's evaluation of past relationships; 3) Emotional Investment in Relationships (EIR) relates to one's capacity for intimacy and emotional sharing; 4) Emotional Investment in Moral Standards (EIM) broadly reflects one's ability to think about moral questions and show genuine compassion towards others; 5) Understanding of Social Causality (CS) assesses the extent to which one understands human behavior, or why people act the way they do in various situations; 6) Experience and Management of Aggressive Impulses (AGG) reflect one's ability to tolerate and appropriately express anger; 7) Self-Esteem (SE) assesses one's self esteem; and 8) Identity and Coherence of Self (ICS) assesses one's level of integration versus fragmentation. The SCORS has shown good to excellent reliability when used to rate semi-structured interview data, TAT narratives, early memories narratives, dream narratives, and other clinical data such as psychotherapy session material (cf. Stein, Hilsenroth, Slavin-Mulford, & Pinsker, 2011). For this study we averaged all 8 items of the SCORS-G. Past research (Defife et al., 2013) has used this method. The SCORS-G ratings were based on the clinicians' interactions with the patient both in therapy and on the unit. All available information including background information and behavioral observations of the patient made by the staff were used in the clinicians' SCORS-G ratings. Given that raters scores were averaged together, we calculated the Intraclass Correlation Coefficient (ICC; Shrout & Fleiss, 1979)

using a one-way random approach to assess reliability. ICCs of $< .40$ are considered poor, $.40 - .75$ are considered fair to good, and $> .75$ is considered excellent (Fleiss, 1986). Raters in the present study obtained an ICC ($_{1,2}$) of .65 which is in the good range.

Global Assessment of Functioning (GAF; American Psychiatric Association, 2000): The GAF is a 100-point clinician-rated scale created to assess a patient's overall level of functioning. The scale is behaviorally anchored to help guide clinical ratings. Both the patient's individual and group therapist completed the GAF blind to each others' ratings at admission. Every therapist provided ratings based on all available data about the patients' functioning when admitted and again at discharge. We calculated the ICCs in much the same fashion as for the SCORS-G ratings using a one-way random approach to assess reliability. ICCs of $< .40$ are considered poor, fair = $.40 - .59$, good = $.60 - .74$, and $> .75$ is considered excellent (Shrout & Fleiss, 1979). Raters in the present study obtained an ICC ($_{1,2}$) of .58 for the admission GAF which falls into the fair to good range. Raters in the present study obtained an ICC ($_{1,2}$) of .58 for the admission GAF which falls into the fair range.

Clinician Therapy Engagement Rating: The patients' individual and group psychotherapist rated the overall amount of engagement and participation the patient showed in the modality they treated them (i.e., group or individual therapy) during the length of their hospitalization. They rated them on a scale from 1 ("not engaged") to 5 ("very engaged"). This rating was done blind to the other clinicians' ratings of each patient as well as self-report questionnaire results. These ratings were also completed at the patient's discharge. Raters were instructed to rate the patients overall averaged level of engagement. Training for this included reviewing patients known to the raters but not part of the study to provide guidance on scoring this measure.

History of Eating Disordered Behavior and Nonsuicidal Self-Injury (NSSI): A member of the research staff who was not part of the clinical team reviewed the patients' charts for a history of eating disordered behavior (i.e., bingeing, purging, restricting) and rated it 1 if the patient has no history of eating disordered behavior and 2 if they did. The same rating was used for a presence or absence of nonsuicidal self-harming behavior where 1 was assigned if the patient did not have a history of NSSI and 2 if there was.

The Schwartz Outcome Scale-10 (SOS-10; Blais et al., 1999) is a 10-item Likert-style self-report measuring psychological well-being and distress. Each item is scored on a scale from 0–6 yielding possible total scores from 0–60. Higher scores on the SOS-10 are representative of better psychological health and well-being. Research (Dragomirecka et al., 2006; Haggerty et al., 2010, in press; Laux & Ahern, 2003; Rivas-Vazquez et al., 2001; Young et al., 2003) has shown that the SOS-10 is a valid and reliable measure of quality of life and psychological well-being with adults and adolescents (Haggerty, Walsh, Kahoud, Forlenza, Ahmed & Ditkowsky, in press). Cronbach's alpha in the present study was .92.

Brief Symptom Measure-25 (BSM-25; Blais et al., 2013) is a unidimensional 25-item self-report of severity of psychiatric symptoms. The items generated were guided primarily by the diagnostic categories of the DSM-IV (APA, 2000). The measure was not designed to be a diagnostic tool (e.g., not intended to diagnose Schizophrenia). The measure is scored by

summing up the items, which are scored on a 7-point Likert scale. Previous research has shown that the BSM-25 demonstrates good sensitivity to change and good psychometric properties in adult inpatient populations (Blais et al., 2013) as well as in adolescent inpatients (Haggerty, Kahoud, Walsh, Ahmed, & Blais, 2013). The measure has also shown high correlations with the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983) and the short form of the SCL-90 (Derogatis & Melisaratos, 1983). Cronbach's alpha in the present study was .93.

Results

Table 1 includes the means and standard deviations of all study measures. We ran correlations between this Overall Functioning Scale and the three ratings that made up this composite (*chronic level of personality functioning, school functioning and peer functioning*) and found that all three ratings correlated above .74 with the OFS. These results show support that the OFS is a unitary measure. We performed Pearson product moment correlations to investigate our study hypotheses. Table 2 contains the results of these correlations. In order to control for rater bias we compared the OFS, which was completed by the individual therapist, to the group. Our results show that OFS is significantly negatively correlated to length of stay and history of nonsuicidal self-injury. OFS was significantly positively correlated to group therapist SCORS-G ratings, group therapist ratings of GAF at admission, and global ratings of engagement in individual therapy. The OFS was not related to global ratings of engagement in group therapy or history of eating disordered behavior. In order to investigate the relationship between the OFS and self-reported changes from admission to discharge as measured by the SOS-10 and the BSM-25 we calculated change scores. The change scores for the SOS-10 and the BSM-25 were calculated subtracting the discharge score from the admission score. Results showed that the OFS was positively related to healthy change on both outcome measures.

We performed hierarchical linear regressions to investigate whether the OFS ratings provide incremental validity above the GAF scores in predicting length of stay. We entered the GAF score in the first block and the OFS ratings in the second block. The results showed that the OFS rating provided incremental validity above and beyond GAF scores. We also ran hierarchical linear regressions for the dependent variables in which we entered the OFS score in the first block and GAF in the second block to see if GAF scores added any incremental validity to the OFS scores. Our results showed that GAF did not add any incremental validity above and beyond the OFS scores. The results of these regressions are presented in Tables 3 and 4. These results further reveal that the OFS may provide an overall assessment of global functioning that outperforms the existing GAF ratings in predicting length of stay.

STUDY 2: INTER-RATER RELIABILITY

Participants—The sample consisted of 32 patients, 59% female, consecutively admitted to the Adolescent Psychiatric Inpatient Unit of a large northeastern hospital. Average length of stay for this sample was 12.7 days. Patients were between the ages of 13–17 which a mean age of 16.1 (SD = 1.04). Ethnic makeup of the sample was as follows: 38% Latino/Hispanic/

Spanish, 27% Caucasian, 23% African American, 8% Other, and 4% Asian. The primary diagnosis for these 32 patients were as follows: 56% Mood Disorders, 41% Conduct Disorder/Oppositional Defiant Disorder, 1% Impulse Control. Patients who showed cognitive impairment or IQ below 70 were excluded from this study (none met this criteria).

Procedures—To investigate the inter-rater reliability of the OFS we had two raters complete the OFS blind to each others' ratings on 32 consecutively admitted patients to the same unit. The OFS ratings were completed within the first few days after admission as in study 1. These 32 patients were not patients in study 1.

Setting—The setting was the same as in study 1.

Measure

Overall Functioning Scale (OFS): The OFS was created from three ratings of global functioning on the Clinician Data Form-Adolescent Version (CDF-A; Westen, Shedler, Durrett, Glass, & Martens, 2003). The CDF-A is used to gather a wide range of demographic and diagnostic information about the patient. For this study we investigated the clinical validity of three of the items which comprise overall functioning in adolescents which include: Chronic level of personality functioning (1 = severe personality disorder to 5 = high functioning); quality of peer relationships (1 = very poor or absent to 5 = very good); and school functioning (1 = failing/dropped out to 5 = working to full potential). These ratings were completed by the patient's individual therapist based on all available information. The OFS was completed after the initial family meeting and psychiatric interview with the patient. All of which happen within the first two business days after admission.

Results

We had 2 clinicians blindly complete the OFS on 32 consecutively admitted adolescent patients admitted to the same service. This sample of patients was not significantly different from the study sample on any of the demographic information (i.e., gender, age, primary diagnosis, ethnicity). We calculated the Intraclass Correlation Coefficient (ICC; Shrout & Fleiss, 1979) using a one-way random approach to assess reliability which yielded an overall ICC ($_{1,2}$) of .88. The ICCs ($_{1,2}$) for the three items of the OFS were .88 for Chronic Personality Functioning, .76 for Peer Relationships, and .85 for School Functioning which all fell above the "excellent" range (Shrout & Fleiss, 1979).

Discussion

Assessing functioning for adolescents is important not only to evaluate whether a particular individual requires treatment but also to assist in identifying if a particular intervention has efficacy. Ultimately, the goal of treatment is to improve overall functioning. Given the climate of evidenced-based medicine, assessing broader concepts beyond the patient's symptoms has become important. The changes to DSM-5 have removed Axis V and the multiaxial diagnosis. As a result, clinicians may not be assessing, in a quantifiable way, the level of functioning of their adolescent patients. This paper investigated the validity and reliability of a brief measure of overall functioning, Overall Functioning Scale. This

assessment can communicate, in broad strokes, how their patient is doing overall based on three core domains of functioning; chronic/general personality functioning, quality of peer relationships, and school performance.

Although measures exist to assess adolescent patients' overall functioning, the OFS looks to address some of the shortcomings of these existing measures. Like the CGAS and GAF, the OFS provides a quantifiable overall rating of the adolescent's functioning. However, although the CGAS and GAF ratings take into consideration functioning on several domains, the clinician is only able to supply a single numerical to quantify functioning. The OFS yields an overall score and clinicians can also see if there is a discrepancy in functioning across the three domains. For example, some patients may show lower functioning in interpersonal relationships but perform well in school. In addition, while the GAF demonstrates only fair reliability in clinical settings the OFS demonstrates excellent reliability with minimal training. This is especially important if a measure is to be used clinically as additional reliability training can be time and money consuming. The Columbia Impairment Scale (Bird, 1999) and the Brief Impairment Scale (Bird, unpublished) also assess overall functioning but are 13 and 23 items long respectively. As a result, these measures may not be as readily used on fast-paced brief stay acute inpatient units where clinician fatigue with completing outcome measures may be an obstacle.

Our results show that the OFS is related in theoretically meaningful ways to variables measuring social cognition and object relation abilities, global functioning as measured by the GAF, global ratings of engagement in individual therapy, history of nonsuicidal self-injury and change on self-reported treatment outcome measures. The results also demonstrated that adolescent inpatients can be reliably rated by clinicians based on the clinicians' unique experience with the patients. Anecdotally, clinicians who completed the OFS ratings both in the initial sample and also in the reliability sample reported that they found the measure easy to use and clinically useful.

The OFS was not found to be related to either global ratings of engagement in group therapy or patients' history of eating disordered behavior. Perhaps patients who had a history of eating disordered behavior demonstrated varied levels of overall functioning. Also, unlike patients with NSSI, patients with a history of eating disordered behavior were not currently demonstrating this behavior. Patients demonstrating this type of behavior are often treated at different unit, either pediatric or specialized eating disordered unit. The fact that the OFS was not related to global ratings of engagement in group therapy was puzzling considering that the OFS was related to engagement ratings for individual therapy. One possible explanation might be that lower functioning adolescents function better in group settings in which they are surrounded by other adolescents who have similar problems. They may engage more when they feel that others have the same or more severe issues. Future research would have to look closer at this finding to understand this relationship.

The current results show support for the OFS to be used as a measure of functioning.

Several limitations of the present study needed to be highlighted. Although this study used the OFS in everyday clinical practice during the course of this study we could not collect

inter-rater reliability data for the research sample. We have no way of knowing if the clinicians who completed this assessment in our initial sample did so in a reliable manner. However, we did explore the reliability of the measure in a second smaller sample and found the reliability of the OFS to be in the “excellent” range overall (Fleiss, 1986). It stands to reason that if the measure shows substantial reliability that the measure was reliably completed in our initial sample but we can never be sure. Future research should investigate the inter-rater reliability in conjunction to the construct validity of the measure.

This study used a mixed model of assessment where clinician, patient-rated measures, and behavioral history from the chart were used. That being said, many of the measures used to demonstrate construct validity (with the exception of outcome, NSSI and eating disordered behavior) were clinician-rated measures. Rater overlap occurred on several measures. The same clinician who was one of the raters for the admission GAF score was also either the clinician who provided the ratings of engagement in individual or group therapy. This could result in inflated correlations between the study measures. Future research would need to have outside observers who were blind to the other study variables rate the level of engagement in therapy.

Another limitation is that the OFS ratings were all completed by an inpatient team member. Although the clinician who completed the ratings used all available collateral data on the adolescent, future research should investigate any similarities and differences by having the adolescent’s family members and teachers complete the ratings as well. Because of this, we are also unable to determine if laypersons without any clinical training would be able to reliably complete this measure.

As with any measure of functioning, what one culture perceives as a mild functional limitation another could view as more severe. We did not specifically investigate cultural aspects to the ratings and did not find any significant differences in scoring based on gender or ethnic background. Our sample size might have not been large enough to investigate this cultural aspect adequately. Future research should investigate whether there are cultural differences in ratings given on this measure with a larger sample.

We also were not able to gather either test-retest reliabilities for the OFS or look at the measures’ sensitivity to change. Both of which are needed in order for the measure to be used for outcome purposes. Future research needs to investigate these aspects of the OFS with larger clinical samples as well as perhaps with outpatient samples.

These limitations aside, the Overall Functioning Scale shows promise as a measure to assess functioning with adolescents on an inpatient service. The measure demonstrates good construct validity in that it is related to other constructs in theoretically meaningful ways and also demonstrates substantial inter-rater reliability. The OFS was also related to length of stay which as stated above has become a metric in which service is measured on inpatients units.

Acknowledgments

The preparation of the manuscript was supported in part by a grant awarded to the first author from the American Psychoanalytic Association and by National Institute of Mental Health (NIMH) Grant No. 1R21MH097781-01A1 received by the first author.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-Text Revision. American Psychiatric Pub Inc; 2000.
- Bird, HR. The assessment of functional impairment. In: Shaffer, D.; Lucas, CP.; Richters, JE., editors. Diagnostic assessment in child and adolescent psychopathology. New York: Guilford Press; 1999. p. 209-229.
- Bird HR, Canino GJ, Davies M, Ramírez R, Chávez L, Duarte C, Shen S. The Brief Impairment Scale (BIS): A multidimensional scale of functional impairment for children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2005; 44(7):699–707. [PubMed: 15968239]
- Bird HR, Canino G, Rubio-Stipec M, Ribera JC. Properties of the Children's Global Assessment Scale. *Archives of General Psychiatry*. 1987; 44:821–824. [PubMed: 3632256]
- Blais M, Blagys MD, Rivas-Vazquez R, Bello I, Sinclair SJ. Development and Initial Validation of a Brief Symptom Measure. *Clinical Psychology & Psychotherapy*. 2013
- Blais MA, Lenderking WR, Baer L, deLorell A, Peets K, Leahy L, Burns C. Development and initial validation of a brief mental health outcome measure. *Journal of Personality Assessment*. 1999; 73(3):359–373. [PubMed: 10689649]
- Blais MA, Matthews J, Lipkis-Orlando, Lechner E, Jacobo M, Lincoln R, Gulliver C, Herman JB, Goodman AF. Predicting length of stay on an acute care medical psychiatric inpatient service. *Administration and Policy in Mental Health and Mental Health Service Research*. 2003; 31(1):15–29.
- Compton MT, Craw J, Rudisch BE. Determinants of inpatient psychiatric length of stay in an urban county hospital. *Psychiatric Quarterly*. 2006; 77(2):173–188. [PubMed: 16763767]
- DeFife JA, Goldberg M, Westen D. Dimensional assessment of self- and interpersonal functioning in adolescents: implications for DSM-5's general definition of personality disorder. *Journal of Personality Disorders*. (in press).
- Derogatis L, Melisaratos N. The Brief Symptom Inventory: A brief introductory report. *Psychological Medicine*. 1983; 13(3):595–605. [PubMed: 6622612]
- Epstein, MH.; Kutash, K.; Duchnowski, AJ. Outcomes for children and youth with emotional and behavioral disorders and their families: Programs and evaluations best practices. 2. Pro-Ed; 2005.
- Fleiss, JL. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986.
- Gold J, Shera D, Clarkson B. Private psychiatric hospitalization of children: Predictors of length of stay. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1992; 32(1):135–143. [PubMed: 8428866]
- Gordon RE, Jardioli P, Gordon KK. Predicting length of stay of psychiatric patients. *American Journal of Psychiatry*. 1985; 142:235–237. [PubMed: 3918469]
- Horowitz L, Inouye D, Siegelman E. On averaging judges' ratings to increase their correlation with an external criterion. *Journal of Consulting and Clinical Psychology*. 1979; 47(3):453–458. [PubMed: 528713]
- Kazdin AE. Psychotherapy for children and adolescents. *Annual Review of Psychology*. 2003; 54:253–276.
- Kuurmann K, Hilsenroth MJ. Exploring in-session focus on the patient-therapist relationship: Patient characteristics, process and outcome. *Clinical Psychology and Psychotherapy*. 2012; 19:187–202. [PubMed: 21370317]
- Peters EJ, Hilsenroth MJ, Eudell-Simmons EM, Blagys MD, Handler L. Reliability and validity of the Social Cognition and Object Relations Scale in clinical use. *Psychotherapy Research*. 2005; 16(5): 617–626.

- Roedelof AJM, Bongers IL. Treatment engagement in adolescents with severe psychiatric problems: A latent class analysis. *European Child and Adolescent Psychiatry*. 2013; 22:491–500. [PubMed: 23455601]
- Selby EA, Bender TW, Gordon KH, Nock MK, Joiner TE. Non-suicidal self-injury (NSSI) Disorder: A preliminary study. *Personality Disorders: Theory, Research, and Treatment*. 2012; 3(1):167–175.
- Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S. A children's global assessment scale (CGAS). *Archives of General Psychiatry*. 1983; 40(11):1228–1231. [PubMed: 6639293]
- Shedler J, Karliner R, Katz E. Cloning the clinician: A method for assessing illusory mental health. *Journal of Clinical Psychology*. 2003; 59(6):635–650. [PubMed: 12754694]
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86(2):420–428. [PubMed: 18839484]
- Stein MB, Hilsenroth MJ, Pinsker-Aspen JH, Primavera L. Validity of DSM-IV Axis V Global Assessment of Relational Functioning Scale. *Journal of Nervous and Mental Disease*. 2009; 197:50–55. [PubMed: 19155810]
- Stein, M.; Hilsenroth, M.; Slavín-Mulford, J.; Pinsker, J. Social Cognition and Object Relations Scale: Global Rating Method (SCORS-G). Boston, MA: Massachusetts General Hospital and Harvard Medical School; 2011. unpublished manuscript Retrieved January 8, 2013, from www.psychsystems.net/Manuals
- Weiss B. Routine monitoring of the effectiveness of child psychotherapy. *Journal of Child Psychology and Psychiatry*. 1998; 39(7):943–950. [PubMed: 9804027]
- Wentz E, Gillberg C, Anckarsater H, Gillberg C, Rastam M. Adolescent-onset anorexia nervosa: 18-year outcome. *The British Journal of Psychiatry*. 2009; 194:168–174. [PubMed: 19182181]
- Westen D. A clinical-empirical model of personality: Life after the Mischel era and the NEO-lithic era. *Journal of Personality*. 1995; 63:495–524. [PubMed: 7562363]
- Westen D, Arkowitz-Western L. Limitations of the Axis II in diagnosing personality pathology in clinical practice. *American Journal of Psychiatry*. 1998; 155:1767–1771. [PubMed: 9842791]
- Westen D, Shedler J, Durrett C, Glass S, Martens A. Personality diagnoses in adolescence: DSM-IV axis II diagnoses and an empirically derived alternative. *American Journal of Psychiatry*. 2003; 160(5):952–966. [PubMed: 12727701]
- Williams, R.; Kerfoot, M. Strategy, planning, delivery and evaluation. Oxford University Press; 2005. Child and adolescent mental health services.
- Winters NC, Collett BR, Myers KM. Ten-year review of rating scales, VII: Scales assessing functional impairment. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2005; 44(4):309–338. [PubMed: 15782079]
- Zittel-Conklin C, Westen D. Borderline personality disorder in clinical practice. *American Journal of Psychiatry*. 2005; 162:867–875. [PubMed: 15863787]

Appendix. A

Overall Functioning Scale

1. Chronic level of personality functioning (*circle number*):

2. Quality of peer relationships (*circle number*):

3. School functioning (*circle number*):

Figure 1.
Scree plot of OFS principal component analysis

Table 1

Means and standard deviations for study measures

	N	Mean	SD
Overall Functioning Scale	66	7.91	1.79
Group therapists' SCORS-G rating	62	3.17	.63
Group therapists GAF score at admission	66	40.86	7.68
Global Rating of Engagement in Group Therapy	65	2.71	1.11
Global Rating of Engagement in Individual Therapy	57	3.28	1.19
SOS-10 at admission	70	36.76	14.09
SOS-10 at discharge	58	44.34	11.98
BSM-25 at admission	68	50.84	30.47
BSM-25 at discharge	58	35.71	27.93

Note: SCORS-G = Social Cognition and Object Relations Scale- Global Version; GAF = Global Assessment of Functioning; SOS-10 = Schwartz Outcome Scale-10; BSM-25 = Brief Symptom Measure-25

Table 2

Means and Standard Deviation of Overall Functioning Scale and adjusted item-scale correlations given at admission.

OFS Item	Mean	Standard Deviation	Item-Scale Correlation
Chronic Level of Personality Functioning	2.7	.65	.54
Peer Functioning	2.7	.78	.42
School Functioning	2.48	.90	.41

Table 3

Pearson product correlations between Overall Functioning Scale ratings and study measures.

	<i>N</i>	OFS	
		<i>r</i>	<i>p</i>
Length of stay in days	59	-.39	.002
SCORS-G [*]	59	.34	.01
GAF at admission [*]	61	.30	.02
Global Rating of Engagement in Group Therapy	60	.03	<i>ns</i>
Global Rating of Engagement in Individual Therapy	54	.33	.02
History of NSSI	66	-.27	.03
History of Eating Disorder Behavior	66	-.13	<i>ns</i>
SOS-10 ^t	50	.35	.01
BSM-25 ^t	49	-.46	<.001

Note:

^{*} OFS was completed by the individual therapist. We compared the OFS with the group therapists' ratings on these items.

^t change scores were calculated by subtracting the discharge score from the admission score.

Table 4
Summary of Hierarchical Regression Analysis for Variables Predicting Length of Stay

Variable	Model 1			Model 2		
	B	SE B	β	B	SE B	β
GAF	-.10	.10	-.14	.03	.02	.17
OFS				.60	.23	.37
R^2		.08*			.21**	
F for change in R^2		4.60*			9.08**	

* $p < .05$.

** $p < .01$.