Reliability and Validity of the Self-Assessment of Occupational Functioning

Alexis D. Henry, Kathi Brenneman Baron, Laurie Mouradian, Clare Curtin

Key Words: evaluation process, occupational therapy

Objective. Two studies examined the reliability and validity of the Self-Assessment of Occupational Functioning (SAOF), a 23-item self-assessment of perceptions of strengths and weaknesses relative to occupational functioning, grounded in the Model of Human Occupation.

Method. The first study examined the test–retest reliability of the SAOF and involved 37 college students without disabilities who completed the SAOF twice. The second study, which involved 39 young persons hospitalized with psychiatric disorders, examined internal consistency reliability of the SAOF and examined correlations between SAOF scores and composite scores on the Self-Perception Profile, a widely used measure of perceived competence. In addition, data from both studies were combined to examine the ability of the SAOF to discriminate between the college students without disabilities and the young persons with psychiatric disorders.

Results. Kappa and intraclass correlation coefficients (ICCs) were used to examine test–retest reliability and Cronbach’s alpha was used to examine internal consistency. Acceptable levels of test–retest (ICCs) and internal consistency (Cronbach’s alpha) reliability were found for the subscale and total scores of the SAOF. However, test–retest reliability (kappa) was lower than desirable for many of the individual SAOF items. The young persons with psychiatric disorders had lower item, subscale, and total scores on the SAOF than did the college students without disabilities. In addition, a discriminant analysis predicting group membership (college students without disability vs. young persons with psychiatric disorder) correctly classified 76.6% of the participants based on the four subscale scores of the SAOF.

Conclusion. The SAOF has the potential to be a reliable and valid clinical assessment; however, additional research is needed.


Ideally, occupational therapy evaluation allows a practitioner to achieve an understanding of a client in a manner that enhances the overall effectiveness of the therapeutic relationship and intervention. In order to be effective, evaluation should actively involve the client in the identification of problems for occupational therapy intervention. A client-centered approach to evaluation, where the occupational therapist and client participate together in the identification of problems and priorities for intervention, has been advocated by many in the field (Law et al., 1994; Mattingly, 1991; Neistadt, 1995). This shared understanding is needed for the therapist and client to jointly set goals and move forward in their work together (Mattingly, 1991). In addition to being an interactive process between therapist and client, Kielhofner and Mallinson (1995) argued that gathering assessment data is also an analytic process because it is “theoretically
informed...[by] a framework provided by the model(s) of practice a therapist is using” (p. 189).

The Self-Assessment of Occupational Functioning (SAOF) (adult version) (Baron & Curtin, 1990) is a theoretically informed, interactive assessment and goal-setting procedure. Grounded in the Model of Human Occupation (MOHO; Kielhofner, 1985, 1995), the SAOF has undergone several revisions since it was originally developed in 1985. Both child and adult versions of the SAOF have been developed. The adult version of the SAOF includes a 23-item checklist-format measure of a person's perception of his or her strengths and weaknesses relative to occupational functioning. Nine items tap aspects of the volition subsystem, six items represent aspects of the habituation subsystem, seven items represent aspects of the performance subsystem, and one item taps the environment. Each item is rated using a 3-point ordinal scale (3 indicates that the area is a personal strength; 2 indicates an adequate area of function; 1 indicates that the area needs improvement). After completing the checklist, the client identifies the three or four items that are a priority for change in his or her life. The SAOF data provide the jumping-off point for collaborative goal-setting between the therapist and client. A manual providing instruction for the administration of the SAOF and the ensuing collaborative goal setting has been developed (Baron & Curtin, 1990).

Such tools have both clinical and research use. The SAOF has been the most frequently requested assessment from the Model of Human Occupation Clearinghouse at the University of Illinois at Chicago (G. Kielhofner, personal communication, April 1998), suggesting that therapists from around the world find the tool clinically useful. Whereas early unpublished studies suggested that the SAOF is content valid (Baron, 1986), data on the psychometrics of the assessment have not been available. In this article, we present the results of two studies that examined the reliability and validity of a preliminary 27-item version of the SAOF. (Four volition subscale items were dropped from the 27-item version to create the 23-item version; otherwise, items on the two versions are identical. The preliminary version used two options for rating "[strength" or "needs improvement"] rather than the three options of the 23-item version).

The first study examined test–retest reliability of the SAOF among college students without disabilities. The second study examined internal consistency reliability and concurrent validity of the SAOF among adolescent and young adults with psychotic disorders. In addition, data from these two studies were combined in order to examine whether the SAOF discriminated between young persons with and without psychiatric disorders.

Study 1: Test–Retest Reliability of the SAOF

Participants

Thirty-seven undergraduate and graduate occupational therapy students participated in the test–retest reliability study. The participants included 36 women and 1 man, ranging in age from 20 to 45 years (M = 24.98, SD = 6.58).

Procedure

Consenting participants were administered the 27-item adult version SAOF on two separate occasions, 2 weeks apart. During each administration, participants completed the SAOF in a quiet classroom, with only the participant and an examiner present. Using an ordinal scale, responses on individual items of the SAOF were scored as 1 (strength) or 0 (needs improvement). The mean scores on items comprising each subscale were used for the four subscale scores (volition, habituation, performance, and environment); the total score was the mean of all 27 items. Agreement between the first and second administrations of the SAOF was examined using kappa and intraclass correlation coefficients (ICCs). Kappa coefficients (Cohen, 1960) were calculated for each of the 27 items, and ICCs (2, 1) (Shrout & Fleiss, 1979) were calculated for the four subscale scores and the total score.

Results

The test–retest reliability coefficients are presented in Table 1. The kappa coefficients for the 27 items ranged from -.06 to 1.00. The ICCs were .70 for the volition subscale, .74 for the habituation subscale, .74 for the performance subscale, and .68 for the environment subscale; the coefficient for the total score was .87.

Study 2: Internal Consistency and Validity of the SAOF

Internal consistency and concurrent validity of the SAOF were examined as part of a larger prospective follow-up study of adolescents and young adults with a first episode of psychotic symptoms (Henry, 1994). Concurrent validity was examined by comparing scores on the SAOF to scores on an established measure of perceived competence, the Self-Perception Profile (Messer & Harter, 1986; Neemann & Harter, 1986).

Participants

Thirty-nine adolescents and young adults, recruited during their first psychiatric hospitalization, participated in the study. The participants included 17 women and 22 men, ranging in age from 17 to 25 years (M = 21.24, SD = 2.08). Thirteen (33.3%) participants were diagnosed with schizophrenia, schizoaffective disorder, or another psychotic disorder, and 26 (66.6%) were diagnosed with bipolar disorder or major depression. Twenty-nine (74.4%) of the participants were college students immediately before hospitalization; the rest were employed.

Procedure and Instruments

The data collection procedures for the prospective study have been described in detail elsewhere (Henry, 1994). As
part of the prospective study, participants completed several self-report measures during hospitalization, including the adult version of the SAOF and an age-appropriate version of the Self-Perception Profile (college student or adult version) (Messer & Harter, 1986; Neemann & Harter, 1986). The measures were administered individually, and participants were given verbal and written instructions for completing each measure by the principal investigator (AH). The SAOF was scored in the same manner as described in Study 1.

The Self-Perception Profiles (SPP). The SPP measure perceptions of competence along several dimensions. The college student version of the SPP taps competency perception in 13 age-appropriate domains, and the adult version taps 12 domains. The specific domains tapped vary slightly across versions; examples include scholastic competence, social acceptance, job competence, and global self-worth. Both versions use the same 4-point ordinal scale for rating items; a higher score reflects higher perceived competence. The SPP authors (Messer & Harter, 1986; Neemann & Harter, 1986) reported internal consistency reliability coefficients for the college student subscales ranging from .76 to .92, and for the adult subscales ranging from .63 to .92.

The SPP subscales scores are not typically summed for a total score; thus, the two versions yield 12 or 13 scores. However, in order to reduce the overall number of variables in the data set used in the prospective study, certain SPP subscales were collapsed into composites using principal components analyses (PCA). All subscale scores (except global self-worth) for the college student and adult versions were separately subjected to the PCA. Based on the results, two composite variables were created. Productive role competence represents competency perceptions such as job competence, scholastic competence, and household management; social competence represents competency perceptions such as sociability, romantic relationships, friendships, and humor. Internal consistency coefficients (Cronbach's alpha) for productive role competence were .72 for the college student version and .65 for the adult version; for social competence, the coefficients were .58 and .83. Because Harter and colleagues argue that global self-worth is a construct distinct from beliefs regarding competencies in life domains (Harter, 1990), scores on the global self-worth subscales were retained as a separate variable.

**Data Analysis**

Relationships among the four subscale and total scores on the SAOF and the productive role competence, social competence, and global self-worth scores on the SPP were examined using Spearman rank order correlation coefficients. In addition, internal consistency estimates were calculated for the volition, habituation, and performance subscales and for the total score of the SAOF using Cronbach's alpha. Because it is composed of a single item, internal consistency cannot be calculated for the environment subscale.
Results

Table 2 presents Spearman correlations between the SAOF subscale and total scores and the SPP scores. Correlations between the SPP scores and the SAOF total score, and volition, habituation, and performance subscale scores ranged from .05 to .61; most were moderate and all but two were statistically significant. The correlations between the SPP scores and the environment subscale were low and not significant. Internal consistency estimates were .88 for the total score, and .83 for the volition subscales, .70 for the habituation subscales, and .66 for performance subscales.

Combined Data From Studies 1 and 2

Method

Data from the first administration of the SAOF in Study 1 were combined with the SAOF data in Study 2 in order to examine similarities and differences between the two groups of participants (i.e., typical young persons [college students] and young persons with psychiatric disorders) on the SAOF, and to examine the extent to which the SAOF discriminated between the two groups. The two groups were significantly different in age ($t = -3.23, df = 42.8, p = .003$); the typical college students were older.

Results

Because the SAOF items were scored 0 (needs improvement) or 1 (a strength), mean item scores reflect the proportion of participants who indicated that the item was a strength. For both groups of participants, individual scores ranged from 0 to 1 on all items. However, on every item, the participants with psychiatric disorders had lower mean scores than the college students, and these differences were statistically significant for 12 of the items. Table 3 shows the five items with the lowest mean scores, and the five items with the highest mean scores for both groups. Although the hospitalized participants consistently had lower mean scores, the five items most often identified as “needs improvement” were the same for the two groups.

Table 4 shows means, standard deviations, minimum and maximum scores, and skewness for the volition, habituation, performance, and environment subscales and for the total score for two groups. Because the college students’ mean scores were quite high for many items, their mean subscale scores were skewed in comparison to the hospitalized participants. (The closer a skewness value is to 0, the more normally distributed the score.) This was particularly true for the performance and environment subscales.

Analysis of covariance (controlling for age) was used to examine differences between the groups on the four subscale and total scores. With age controlled for, mean subscale and total scores for the young persons with psychiatric disorders were significantly lower than those for the college students.

Finally, a stepwise discriminant analysis, predicting group membership (i.e., typical college students vs. young persons with psychiatric disorders), was performed using the four SAOF subscale scores as predictor variables. The discriminant analysis also identifies which subscale contributes most to the prediction of group membership. The analysis correctly classified 76.6% of the subjects based on the four predictor variables. Among those misclassified, 13 persons with psychiatric disorders were classified into the group without disabilities, and five college students were classified into the group with psychiatric disorders. Of the four subscales, the performance subscale contributed most to the prediction of group membership ($F = 17.74, df = 1, 74, p = .0001$). With the performance subscale accounted for, the environment subscale was a near significant predictor ($F = 3.84, df = 1, 73, p = .054$). The volition and habituation subscales were not significant predictors in the discriminant analysis.

Discussion

Reliability and validity studies such as these serve to increase clinicians’ understanding of the dependability and meaningfulness of clinical assessment data. Internal consistency reliability coefficients give an indication of the extent to which the items are intercorrelated; that is, the extent to which the items are measuring the same phenomenon. Test–retest reliability coefficients give clinicians an indication of the extent to which a person’s scores on a measure would be expected to be stable, at least over the time frame used in the study. There are no absolute standards for “acceptable” reliability, although some authors have suggested that coefficients of .60 or higher are acceptable for test–retest reliability, and coefficients in the range of .70 to .80 are acceptable for internal consistency (Benson & Clark, 1982; Nunnally, 1978). Of course, the higher the coefficient (the closer to 1.00), the better. In practical terms, “acceptable” reliability is dependent, to a large extent, on the purpose of the measure and the degree of

Table 2

Spearman Correlations Between Subscale Scores on Self-Assessment of Occupational Functioning (SAOF) and Scores on Self-Perception Profile (SPP) Composites Among Young Persons With Psychiatric Disorders

<table>
<thead>
<tr>
<th>SSP</th>
<th>Volition</th>
<th>Habitation</th>
<th>SAOF Performance</th>
<th>Environment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productive role competence</td>
<td>.39**</td>
<td>.05</td>
<td>.32*</td>
<td>.060</td>
<td>.29</td>
</tr>
<tr>
<td>Social competence</td>
<td>.53***</td>
<td>.34*</td>
<td>.61***</td>
<td>-.002</td>
<td>.56***</td>
</tr>
<tr>
<td>Global self-worth</td>
<td>.57***</td>
<td>.39**</td>
<td>.52***</td>
<td>.060</td>
<td>.57***</td>
</tr>
</tbody>
</table>

Note. $n = 39$. SAOF was scored: 1 = a strength, 0 = needs improvement. SPP was scored: 1 = low perceived competence, 4 = high perceived competence.

*p < .05, **p < .01, ***p < .001.
error (or unreliability) that is tolerable. For example, a diagnostic test for a serious medical condition should have very high reliability (mid-.90s or higher), especially if decisions about how to proceed with treatment rest on the results of the test. Lower levels of reliability may be tolerable for other types of measures (DeVellis, 1991; Polgar, 1998).

The SAOF is intended to be used to facilitate collaborative goal-setting between the clinician and the client. Because a client’s goals and priorities are likely to change over time, implied in such a procedure is an ongoing reevaluation of the client’s goals and a renegotiation of the therapeutic work. Given the intended use of the SAOF, the four subscale scores and total score of the SAOF show acceptable test–retest reliability. Among the college student participants, the subscale and total scores were fairly stable over time. The ICCs indicate that the college students were generally identifying the same number of volition, habituation, performance, and environment strengths, as well as the same total number of strengths, at each administration of the SAOF. All ICCs for the subscale and total scores met the standard for test–retest reliability suggested by Benson & Clark (1982).

The internal consistency estimates for the volition, habituation, and performance subscales and the total scores approach or exceed the standard of .70 recommended by Nunnally (1978). Although previous content validity studies (Baron, 1986) suggested that the SAOF items adequately represent MOHO constructs, the internal consistency coefficients suggest that each set of subscale items are related to each other and together form a conceptually meaningful construct.

Whereas the subscale and total scores showed acceptable test–retest reliability, the kappa coefficients for many individual items were lower than preferable. Only eight of the items showed substantial agreement (.60 or above). Thirteen items had coefficients in the moderate range (.40 to .59), and six showed low agreement (.39 or below). Individual item reliability coefficients are generally expected to be lower than those for subscale and total scores. Regardless, it is important to consider factors that may account for the low reliability of certain items. Again, the SAOF is intended to be used jointly by clinicians and clients as a means of identifying problems to be addressed by occupational therapy interventions. However, as a group, the college students without disabilities who participated in the test–retest reliability study tended to rate most items as “a strength” and identified very few problems. Thus, many items had very limited variability across the two administrations of the SAOF. A change in response from the first to the second administration among only a few participants on items that otherwise had low variability will result in a very low estimate of agreement. In other words, the low reliability coefficients for the items primarily result from this low variability rather than from an actu-

### Table 3

<table>
<thead>
<tr>
<th>Item</th>
<th>College Students</th>
<th>Young Persons With Psychiatric Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low score items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having a healthy balance of roles</td>
<td>.46 .51</td>
<td>Organizing my time</td>
</tr>
<tr>
<td>Staying with a frustrating activity</td>
<td>.49 .51</td>
<td>Staying with a frustrating activity</td>
</tr>
<tr>
<td>Having habits that support my roles</td>
<td>.51 .51</td>
<td>Having habits that support my roles</td>
</tr>
<tr>
<td>Organizing my time</td>
<td>.57 .51</td>
<td>Having a healthy balance of roles</td>
</tr>
<tr>
<td>Expressing myself to others</td>
<td>.57 .51</td>
<td>Expressing myself to others</td>
</tr>
<tr>
<td>High score items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believing I can make things happen</td>
<td>.31</td>
<td>Believing I can make things happen</td>
</tr>
<tr>
<td>Performing daily living skills</td>
<td>.28</td>
<td>Having future goals</td>
</tr>
<tr>
<td>Being involved in my roles</td>
<td>.23</td>
<td>Identifying my interests</td>
</tr>
<tr>
<td>Having future goals</td>
<td>.23</td>
<td>Doing activities that give</td>
</tr>
<tr>
<td>Being physically able to do</td>
<td>.23</td>
<td>me a sense of purpose</td>
</tr>
<tr>
<td>do what needs to be done</td>
<td></td>
<td>Being physically able to do</td>
</tr>
<tr>
<td></td>
<td></td>
<td>what needs to be done</td>
</tr>
</tbody>
</table>

Note. n = 57 for college students; n = 39 for young persons with psychiatric disorders. College students’ data were taken from the first administration of the SAOF during the test–retest reliability study. SAOF was scored: 1 = a strength, 0 = needs improvement.

### Table 4

<table>
<thead>
<tr>
<th>Item</th>
<th>College Students</th>
<th>Young Persons With Psychiatric Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Volition Subscale</strong>b</td>
<td>.76</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Habituation Subscale</strong>a</td>
<td>.66</td>
<td>.26</td>
</tr>
<tr>
<td><strong>Performance Subscale</strong>a</td>
<td>.78</td>
<td>.19</td>
</tr>
<tr>
<td><strong>Environment Subscale</strong>a</td>
<td>.84</td>
<td>.37</td>
</tr>
<tr>
<td><strong>Total score</strong>a</td>
<td>.75</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note. n = 57 for college students; n = 39 for young persons with psychiatric disorders. College students’ data were taken from the first administration of the SAOF during test–retest reliability study. SAOF was scored: 1 = strength, 0 = needs improvement.

# Means are significantly different at p < .001. *Means are significantly different at p < .005. †Means are significantly different at p < .05.
al lack of agreement between participants’ responses from the first to the second administration. This was the case for most of the individual items that showed very low reliability. However, it is also possible that a lack of clarity of certain items contributed to their low reliability.

In addition to reliability, these studies examined both the concurrent and discriminant validity of the SAOF. The moderate correlations between SAOF subscale and total scores and the composite scores from the concurrently administered SPP, a well-researched measure of perceived competence, suggest that the SAOF and SPP measure related, but not equivalent, constructs, and provide evidence that the SAOF is a valid measure of a person’s sense of his or her abilities. Not surprisingly, the correlations between the SPP composite scores and the environment subscale score of the SAOF, which addresses aspects of the environment as opposed to the person, were not significant.

When SAOF scores of college students without disabilities were compared with those of young persons hospitalized with a psychotic episode, both similarities and differences emerged. Interestingly, the five items most often identified as “needs improvement” were the same for both groups of participants, suggesting that organizing one’s time, expressing oneself to others, or staying with a frustrating activity are challenges that commonly confront young persons.

As would be expected, the hospitalized participants had lower item, subscale, and total scores in comparison to the college students. This difference was not attributable to age differences between the two groups. With age controlled for, subscale and total scores were significantly lower for the young persons with psychiatric disorders. Clearly, these participants perceived themselves to have fewer occupational strengths than the college students without disabilities. The results of the discriminant analysis suggest that the SAOF is a likely to be a measure that can detect differences between groups with and without disabilities or disorders. It is true that almost 25% of participants were misclassified in the discriminant analysis. However, it is important to note that most of the hospitalized participants had been college students immediately before admission (for most of them, this was a few days to a few weeks before data collection) and that this was the first psychiatric hospitalization for all. Thus, in many ways, the young persons who were hospitalized were similar to the college students without disabilities. The fact that the discriminant analysis more often misclassified the hospitalized young persons into the college student group than vice versa supports this supposition. That the discriminant analysis correctly classified more than 75% of the participants suggests that with different comparison groups (e.g., typical, healthy adults; adults with severe and persistent mental illness), the ability of the SAOF to discriminate between persons with and without disabilities or disorders might be greater.

The participants with psychiatric disorders differed from the college students without disabilities on all of the SAOF subscales. However, for the three subscales related to the individual, the differences were greatest on the performance and habituation subscales (Table 3), and the results of the discriminant analysis showed scores on the performance subscale to be the ones that discriminated most between the two groups. The data suggest that, with the onset of psychotic symptoms, one’s sense of one’s ability to meet everyday expectations, which is tapped by the performance and habituation subscales of the SAOF, may be more compromised initially than one’s sense of identity, which is tapped by the volition subscale. Other occupational therapy authors who have considered the experiences of young persons with psychiatric disorders have noted that these persons often have the same goals and interests that they had premorbidly, while they grapple with the impact the disorder will have on their ability to meet daily challenges (Barrows, 1996; Helfrich, Kielhofner, & Mattingly, 1994).

Conclusions and Need for Additional Research

These results suggest that the SAOF has potential to be a reliable and valid clinical assessment. However, several things could be done that would likely improve the overall dependability and clinical usefulness of the assessment. It appears that the SAOF would benefit from a review and revision of the individual items, and this work is currently underway at University of Illinois at Chicago. In particular, as a single item subscale, the environment subscale has limited usefulness. Expansion of this subscale, incorporating current MOHO constructs regarding the environment (Kielhofner, 1995), is warranted. With these revisions, further studies of the reliability of the SAOF will be needed. In particular, reliability studies with participants with a variety of disabilities must be done, especially because these persons are likely to show more variability in their responses to the SAOF items than participants without disabilities.

Moreover, because the SAOF is a procedure that advocates the client’s voice in goal setting and a client-centered approach to occupational therapy clinical practice, the usefulness of the SAOF as a collaborative goal-setting procedure should be examined. Others who have advocated a client-centered approach to practice have argued for the importance of understanding not just the client’s perspectives on his or her problems, but the extent to which those problems are important to the client (e.g., Law et al., 1994). Incorporating a method for identifying those problems that are of greatest importance to the client would enhance the usefulness of the SAOF. Finally, studies examining the reliability and validity of the children’s version of the SAOF need to be conducted.

Acknowledgments

We thank Amy Anhalt, Lawrence Aronowitz, Rachel Beauchamp, Andrea Benson, Lorraine Burgos, David Busker, Ronald Christopher, James Dailey, Erin Day, Melissa Mondschein, and Tania Roelse for their
assistance with data collection. At the time of these studies, Dr. Henry and Dr. Mouradian were supported, in part, by a grant to Sargent College of Allied Health Professions, Boston University, from the U.S. Department of Health and Human Services, Division of Maternal and Child Health Services.

References


