A Review of Instruments for Measuring Functional Recovery in Those Diagnosed With Psychosis

Brent T. Mausbach1,2, Raeanne Moore3, Christopher Bowie4, Veronica Cardenas2, and Thomas L. Patterson2

1Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0680; 2Alliant International University, San Diego, CA; 3Department of Psychology, Queen’s University, Ontario, Canada

The task of judging an individual’s functional recovery is not an easy one for healthcare professionals. Indeed, increasing one’s accuracy in predicting one’s ability to self-maintain would be of great value for determining if functional recovery has or is occurring. The purpose of this review is to examine existing measures for assessing remission/normalization of functional status among people with psychosis. Our review evaluates 8 measures of functional ability encompassing self-report, clinical, and performance-based measures. We elected to utilize a grading system to aid readers in understanding the merit of a scale for use in assessing functional recovery. In this approach, a letter grade (A, B, or C) was assigned to each of 4 domains we deemed important to professionals in electing to use specific assessments: (1) Ease of Administration, (2) Reliability, (3) Validity/Relationship to Real-World Outcomes, and (4) Sensitivity to Change/Use in Clinical Trials. Results indicated that no “gold standard” measure has been developed to date, but performance-based measures appear to have the most evidence for predicting concurrent self-maintenance abilities (eg, residing independently or maintaining work). More research on existing measures is needed, and greater funding for developing new measures of functional recovery is strongly recommended.

Key words: schizophrenia/self-maintenance/functioning/functional capacity

Introduction

An adequate definition of recovery in schizophrenia has long been elusive, with very different definitions occurring in the scientific and consumer communities.1 Indeed, while the scientific community may view recovery as an outcome defined by its emphasis on symptoms and functioning, consumer-focused definitions may define recovery as a process toward achieving, among other things, empowerment, hope, and respect. While clearly both approaches to defining “recovery” have merit, these methods of defining recovery also yield distinct approaches to measuring the construct. To be clear on the purpose of this review, we have approached this review from a scientific perspective, in which recovery includes dimensions described by Harvey and Bellack.2 Specifically, it should include a protracted period (eg, at least 6 months) of minimal symptoms (ie, positive and negative symptoms), “normal” neuropsychiatric functioning, and the ability for people to function independently in the “real world.” In contrast to clinical remission, functional recovery requires that a person be able to perform daily activities that are required for self-maintenance (earning an income and maintaining a residence). While one might think that measurement of symptoms and cognitive status would mirror functional status, Harvey and Bellack (in this issue) review the extant literature which suggests that improvement/remission of symptoms of psychotic illnesses (eg, hallucinations and delusions) is not necessarily linked to improved functioning nor does there appear to be close links between either of these factors and well-being. Therefore, recovery in those domains appears to require not only separate assessments but also different approaches than what would be used for improving functioning. However, even an improved understanding of what recovery involves leaves a difficult task of measuring whether recovery has occurred.

The task of judging an individual’s functional recovery is not an easy one for healthcare professionals. Using clinical judgment alone may not be enough given clinicians are not embedded into the natural environment of those they work with, thereby making it difficult to know how an individual functions in the real world. Clinicians, therefore, rely on additional measures for assessing functional recovery. These additional measures may be used prior to treatment to help develop an intervention plan, during treatment to “shift course” or to determine progress, and at the conclusion of treatment to aid in a discharge or follow-up plan. Indeed, increasing one’s accuracy in predicting one’s ability to self-maintain...
would be of great value for determining if functional recovery has or is occurring. Along these lines, the purpose of this review is to examine existing measures for assessing remission/normalization of functional status among individuals with psychosis.

Methods

Selection of Measures and Inclusion Criteria

To identify candidate measures for inclusion in our review, studies initially had to meet 4 specific criteria. First, the participants in each study had to have a diagnosis of schizophrenia or schizoaffective disorder. Second, the study must have measured some aspect of functioning, including functional recovery, functional outcome, or functional capacity. Third, studies that measured validity or reliability were included. Fourth, studies were included if they examined independence, work, schooling or education, or community functioning. Guided by these criteria, we searched Medline and PsychInfo databases for journal manuscripts published between January, 1990, and September 2008. Limits were set to include only manuscripts published in English and with adult subjects over 18 years of age. From this search, 87 articles were found. We read the abstracts from these articles to identify potential studies for inclusion as well as manuscript citations to identify further manuscripts that may have initially been missed in our initial search. This overall search yielded 8 measures of functioning, for which an additional search was conducted to identify additional manuscripts focusing on these outcome measures.

Development of a Grading System for Measures

We elected to utilize a grading system to aid readers in understanding the merit of a scale for use in assessing functional recovery. In this approach, a letter grade (A, B, or C) was assigned to each of 4 domains we deemed important to professionals in electing to use specific assessments: (1) Ease of Administration, (2) Reliability, (3) Validity/Relationship to Real-World Outcomes, and (4) Sensitivity to Change/Use in Clinical Trials. Details on each of these domains are given below.

Ease of Administration. Of consideration to many professionals is the length of time a measure requires for administration. Specifically, professionals must consider issues of participant burnout, inclusion in research protocols where several other instruments are administered, and/or use in professional settings where staff time and resources must be considered. When considering these factors, a “brief is better” philosophy was selected. Also of consideration is the requirement of props or testing instruments, which may cause added burden in terms of when and where the instrument may be administered (ie, burden of storage/setup/portability). Examining the 8 scales described in this review, we found that the median time to administer these scales was approximately 30 min. Therefore, in this domain, a grade of “A” was assigned to scales that required fewer than 30 min to administer and had few props. A grade of “B” was assigned to scales that required fewer than 30 min to administer or that had more specialized props. A grade of “C” was assigned to tests requiring 30 min or more to administer and required more props.

Reliability. Data evaluated in determining a grade included intraclass correlation coefficient, test-retest reliability, comparisons of people with psychosis to those without (ie, normal controls), and correlations with scales measuring similar constructs (eg, cognitive functioning). Our criteria were as follows: reliability coefficients above .90 were deemed excellent (grade = A), .80–.89 were deemed good (grade = B), .70–.79 adequate (grade = C), and below .70 were questionable (grade = D).

Validity/Relationship to Real-World Outcomes. For this domain, grades of A were assigned to tests that demonstrated significant concurrent associations with at least 2 real-world outcomes (eg, employment, residential independence) and a prospective relationship with at least 1 outcome. A grade of B was assigned to tests showing significant concurrent associations with at least 2 real-world outcomes. A grade of C was assigned to tests that showed significant concurrent associations with 1 real-world outcome. A grade of “N/A” was assigned where published data were “not available,” and therefore, a determination could not be made regarding the test’s association with real-world outcomes.

Use in Clinical Trials/Sensitivity to Change. For a scale to be useful in measuring functional recovery, scores should change as a function of targeted treatment (eg, skills-training interventions), while those not targeted should demonstrate relatively stable (ie, test scores should remain stable/free of learning effects) scores. The standard for assessing change is the randomized clinical trial. In this domain, a grade of A was assigned to instruments that were used in at least 2 randomized studies and that showed significant benefits to the active treatment condition. A grade of B was assigned to instruments used in at least 1 randomized study showing significant benefits to the active treatment condition, and a grade of C was assigned to instruments used 1 or more non-randomized studies only and demonstrated significant benefit to an active treatment. Grades of N/A were assigned to instruments not utilized in clinical studies.

Results

A summary of the scales assessed in this report, along with grades, is found in table 1.
University of California, San Diego (UCSD)
Performance-Based Skills Assessment

The UCSD Performance-Based Skills Assessment (UPSA)² was developed in 2001 as a means of assessing basic everyday living skills in older people with schizophrenia. The UPSA consists of 5 subtests assessing the individual’s ability in the following domains: (a) Planning/Organization (ie, planning a trip to the beach/zoo), (b) Finances (ie, counting change; writing a check), (c) Communication (ie, calling information to request a phone number; calling the doctor to reschedule an appointment), (d) Travel (ie, reading a bus route map), and (e) Household (ie, completing a shopping list; reading a recipe). Participants receive a score ranging from 0 to 20 for each of the 5 domains, with the 5 domain scores summed to create a total score (range = 0–100).

Ease of Administration. The UPSA requires approximately 30 min to administer, requires no specialized qualifications for administration, and can be administered by a suitably trained paraprofessional. The UPSA also contains a number of props including grocery items, maps of bus schedules, and a telephone, among other items.

Reliability. Harvey et al⁴ report excellent reliability statistics for the UPSA. In a sample of 50 participants with schizophrenia or schizoaffective disorder and 20 normal comparison participants, the interrater reliability coefficient was .91 and 2-week test-retest reliability was .93. In another study, the 4-week test-retest reliability was .70 in a study of 176 participants with schizophrenia.⁵ Using data from the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) trial, Green and colleagues⁵ also indicate that the UPSA demonstrated a small effect size ($d = 0.23$) in terms of practice effects.

Validity/Relationship to Real-World Outcomes. Convergent validity of the UPSA is strong. Patterson and colleagues³ compared UPSA scores of 50 older participants with schizophrenia with those of 20 older normal controls. As expected, participants with schizophrenia demonstrated significantly worse performance than the normal controls on all 5 subscales and total scores. In the MATRICS study, scores on the UPSA were significantly correlated with overall cognitive performance, and this correlation was significantly higher than the correlations between cognitive performance and 2 interview-based measures: (a) the Schizophrenia Cognition Rating Scale and (b) the Clinical Global Impression of Cognition in Schizophrenia—Neurocognitive State Composite score. Correlations between the UPSA and cognitive performance were also significantly higher than that between cognitive performance and the Maryland Assessment of Social Competence Effectiveness score.

Several studies have demonstrated that the UPSA has good criterion validity with real-world outcome measures. Twamley and colleagues⁶ examined the association between UPSA performance and level of independence in a sample of 111 middle-aged and older persons with psychosis. Participants were rated on their level of residential independence from 1 (skilled nursing facility) to 5 (living alone in a house or apartment). The Spearman correlation between total UPSA score and level of independence was .48. A later study by Mausbach et al⁷ also examined the usefulness of the UPSA for predicting current residential status. In this study, 434 middle-aged and older adults with schizophrenia or schizoaffective disorder completed the UPSA and provided information on their current living situation. Participants were classified as either independent (residing alone in a house or apartment) or nonindependent (residing in a care facility such as board and care, skilled nursing facility, etc.). Receiver operator characteristic curves were used to establish an UPSA cutoff for maximum prediction of residential status. Results indicated that at a cutoff of 75, the UPSA predicted current living status with 68% accuracy. Further, the UPSA

---

### Table 1. Evaluation of 6 Scales of Functional Recovery in People With Psychosis

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time to Administer (in min)</th>
<th>Ease of Administration</th>
<th>Reliability</th>
<th>Validity/Relationship to Real-world Outcomes</th>
<th>Use in Clinical Trials/ Sensitivity to Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPSA</td>
<td>30</td>
<td>A/B</td>
<td>B</td>
<td>A/B</td>
<td></td>
</tr>
<tr>
<td>UPSA-Brief</td>
<td>10–15</td>
<td>A</td>
<td>N/A</td>
<td>B/C</td>
<td></td>
</tr>
<tr>
<td>SSPA</td>
<td>20</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>B/C</td>
</tr>
<tr>
<td>MMAA</td>
<td>60–70</td>
<td>C</td>
<td>B/C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>TABS</td>
<td>40</td>
<td>C</td>
<td>B</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ILSS-SR</td>
<td>20–30</td>
<td>A/B</td>
<td>D</td>
<td>N/A</td>
<td>A</td>
</tr>
<tr>
<td>GAF</td>
<td>5–15</td>
<td>B</td>
<td>B/C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ILS</td>
<td>45–55</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>

*Note: UPSA, UCSD Performance-Based Skills Assessment; UPSA-Brief, Brief UCSD Performance-Based Skills Assessment; SSPA, Social Skills Performance Assessment; MMAA, Medication Management Ability Assessment; TABS, Test of Adaptive Behavior in Schizophrenia; ILSS-SR, Independent Living Skills Survey—Self-Report form; GAF, Global Assessment of Functioning; Independent Living Scales; N/A, not applicable.*
was significantly better than Mattis Dementia Rating Scale (DRS) for predicting residential independence.

In addition to using the UPSA to predict residential independence, other studies have examined the UPSA’s usefulness for predicting engagement in community responsibility. Cardenas and colleagues assessed 58 middle-aged and older Latinos with schizophrenia for their engagement in various activities such as working for pay, doing volunteer work, and going to school. Results of this study indicated a linear relationship between the number of responsibilities and scores on the UPSA. Specifically, mean UPSA scores for participants who engaged in 0, 1, and 2 or more responsibilities were 44.9, 58.1, and 67.4, respectively.

Using a modified version of the UPSA consisting of 4 of the 5 test domains, Bowie et al conducted 2 studies examining the UPSA’s association with scores on the Specific Level of Functioning (SLOF) Scale, an observer-rated measure of real-world functioning. In the first, case managers assessed 78 participants with schizophrenia or schizoaffective disorder on their functioning in various SLOF domains, including personal care skills, interpersonal skills, community activities, and work skills. Results indicated that UPSA scores were significantly related to interpersonal skills, community activities, and work skills. That is, those who performed better on the UPSA were rated as having significantly better functioning in these domains. Similar results were found in their second study, which included 222 participants with schizophrenia or schizoaffective disorder. Specifically, a significant relationship was found between UPSA performance and Community Activities and Work Skills ratings, such that higher scores on the UPSA were associated with better functioning.

Use in Clinical Trials/Sensitivity to Change. The UPSA has been used as an outcome measure in a number of clinical trials. Most notable was the Functional Adaptation and Skills Training (FAST) study conducted by Patterson and colleagues. The FAST intervention study compared the effects of a behavioral, skills-training intervention specifically designed to improve everyday functioning in people with schizophrenia (FAST) with that of a support-group style intervention. Both interventions spanned 24 weeks (meeting once per week) and were administered in a group format. The FAST intervention taught participants skills in 6 domains: (a) Medication Management, (b) Social Skills, (c) Communication Skills, (d) Organization and Planning, (e) Transportation, and (f) Financial Management. UPSA scores pre- and post-intervention were examined. Results indicated that participants in the behavioral intervention showed significantly greater change in UPSA performance relative to those in the support-group intervention.

Patterson et al conducted a similar pilot study with Hispanic/Latino participants with schizophrenia or schizoaffective disorder. In this study, 29 participants were randomized to receive a culturally tailored skills-training intervention known as Programa de Entrenamiento para el Desarrollo de Aptitudes para Latinos (PEDAL; n = 21) or to a Spanish-language support condition (n = 8). Participants attended 24 classes on a semiweekly basis (120-min each class). Six-month change in UPSA scores were compared for the 2 conditions, with participants in the PEDAL condition significantly outperforming their counterparts in the support condition (P = .001).

A third study, conducted by Granholm and colleagues, randomly assigned 76 middle-aged and older people with chronic schizophrenia to receive 24 weekly sessions of Cognitive Behavioral Social Skills Training (CBSST) or to receive treatment as usual. Participants in the CBSST intervention received training in various modules, including challenging thoughts (particularly those associated with their illness), asking for support (ie, social skills training), and solving problems (eg, using public transportation, taking medications). UPSA scores were assessed both pre- and postintervention. Results indicated that posttreatment UPSA scores were not significantly different for the CBSST and treatment as usual groups (P = .052; Cohen d = 0.48).

Brief UCSD Performance-Based Skills Assessment

The Brief UCSD Performance-Based Skills Assessment (UPSA-Brief) was introduced in 2007 as a time-sensitive alternative to the full version of the UPSA. Via factor analysis techniques, the UPSA-Brief consists of 2 of the original 5 subscales (ie, Finance and Communications subscales) and, therefore, requires only 10–15 min to administer.

Ease of Administration. As with the original UPSA, the UPSA-Brief requires no specialized qualifications for administration and can be administered by a suitably trained paraprofessional. Significant advantages of the UPSA-Brief over its longer counterpart include its shorter “administration time” (ie, 10–15 min) and fewer props required. In the case of the UPSA-Brief, only a telephone and money are required. Therefore, the UPSA-Brief received a higher Ease of Administration grade than the UPSA.

Reliability. Little data exist on the reliability of the UPSA-Brief. Because the subscales were derived from the original UPSA, test-retest and internal consistency are anticipated to be relatively similar. However, given the UPSA-Brief has fewer items than the original UPSA, reliability is anticipated to be lower for the shorter scale.

Validity/Relationship to Real-World Outcomes. Validity for the UPSA-Brief appears to be good. The Pearson correlation between UPSA and UPSA-Brief scores has been reported to be .91. However, because the UPSA-Brief was derived from scores obtained on the original UPSA, this correlation coefficient may be inflated. Data suggest that the UPSA-Brief is correlated with
measures of overall cognitive functioning and educational attainment but less correlated with symptoms of psychosis. In addition, scores on the UPSA-Brief appear to be related to residential independence and level of community responsibility (see below).

Similar to results for the full version of the UPSA, UPSA-Brief scores were significantly associated with concurrent residential status (ie, independent vs nonindependent), with scores of 60 or greater most discriminat-
ing of individuals who were and were not residing independently in the community. In a separate study, Mausbach et al examined the relations between UPSA-Brief scores and level of community responsibility (eg, work for pay, participate in volunteer work). Individuals who engaged in greater numbers of community responsibilities were predicted to have higher UPSA-Brief scores than those engaging in fewer responsibilities. Results partially supported the overall hypothesis. Among community-dwelling participants (ie, those who resided in a house or apartment), those who engaged in more responsibilities had significantly higher UPSA-Brief scores. However, no significant group differences were observed among participants residing in assisted care facilities (eg, board and care, skilled nursing facility). Although unclear why this relationship was not found, the authors concluded that the expectation of residential facilities may not be to promote community responsibility, despite an individual's capacity to do so.

Use in Clinical Trials. Because it has only recently been introduced, the UPSA-Brief has had little use in clinical trials. Mausbach and colleagues compared change on the UPSA-Brief using subjects from the original FAST study. Participants receiving the behavioral skills-training FAST intervention showed significantly greater improvements in functional capacity relative to partic-

Reliability. In their initial development study, Patterson and colleagues examined interrater reliability and 1-week test-retest reliability for the SSPA total score. The intraclass correlation coefficient was .91 and the test-

Validity/Relationship to Real-World Outcomes. Patterson and colleagues compared SSPA scores of middle-aged and older adults with psychosis to those of age-comparable normal controls. Results were as expected, with the schizophrenia group scoring significantly lower than normal controls. Scores on the SSPA were positively correlated with the quality of well-being (QWB) Scale, a self-report measure of quality of life, and inversely correlated with negative symptoms of psychosis, such that those with greater negative symptoms demonstrated significantly worse social skills capacity as measured by the SSPA. Similarly, participants with worse cognitive functioning demonstrated worse social skills performance.

Pratt and colleagues administered the SSPA to 152 middle-aged and older adults with serious mental illness (ie, schizophrenia, schizoaffective disorder, bipolar disorder, and major depressive disorder). Interrater reliability in this sample was extremely high (item characteristic curve = .99), and scores on the SSPA were positively and significantly correlated with cognitive functioning (ie, DRS scores). Interestingly, in this study, participants with affective disorders scored significantly higher on the SSPA than those with schizophrenic disorders. Finally, a recent study demonstrated significant positive relationships between SSPA scores and overall cognitive ability.

Although introduced in 2001, few studies have exam-
in the relationship of the SSPA to real-world outcomes. One study demonstrated that scores on the SSPA were
positively associated with interpersonal behavior, work skills, and community activity participation in a sample of 222 outpatients with schizophrenia.

**Sensitivity to Change in Clinical Trials.** In their original FAST study, Patterson and colleagues compared change in SSPA scores for participants receiving either a Support intervention (N = 116) or a skills-training intervention (N = 124). Change was assessed from baseline to 6-month postbaseline. Participants receiving the skills-training intervention showed significant improvement in SSPA scores relative to those in the support condition. In a similar study, Patterson et al. offered 29 Latinos with schizophrenia culturally adapted versions of the FAST interventions. However, in this pilot intervention, participants in the skills-training intervention (N = 21) did not show significant differences to support participants (N = 8) in SSPA scores at 6, 12, or 18 months postbaseline. This was likely due to a lack of power to detect differences because Cohen d effect sizes favored the skills-training intervention at all 3 follow-up points (range = 0.49–0.75).

**Medication Management Ability Assessment**

The Medication Management Ability Assessment (MMAA) was introduced in 2002 as a role-play test designed to assess prescription medication management ability in older adults with schizophrenia. Participants are presented with 4 prescribed mock medications, and the interviewer describes the medication regimen to the participants. Additionally, each medication bottle is printed with specific instructions including the following: how many pills to take, what time of day to take the medication, and if the medication should be taken with or without food. The medication bottles are then removed for 1 h, and after the delay, participants are represented with the medication bottles and asked to walk the interviewer through their day, giving the interviewer pills as they would take them throughout the day. Participants receive a total score ranging from 0 to 25 for the handling of each medication (including over- and under-medication) and number of pills correct.

**Ease of Administration.** The MMAA requires no specialized qualifications for administration and can be administered by a suitably trained paraprofessional. Props are required and include mock medication pill bottles with “pills” (eg, colored beans). Because of the props and because the MMAA requires a delayed administration component, in which the test is introduced and completed at a later time point (approximately 1 h), the MMAA received a lower Ease of Administration grade.

**Reliability** In their initial development study, Patterson and colleagues examined 1-week test-retest reliability for the MMAA. In this study, test-retest reliability for the MMAA was excellent, with an intraclass coefficient reported as .96. However, test-retest reliability in other studies has yielded mixed results for the MMAA. In the MATRICS study, the 4-week test-retest coefficient was low (r = .48), particularly when compared with other measures assessed by the MATRICS team. The authors speculated that the low test-retest correlation in this study could be due to a ceiling effect. For practice effects, these researchers also reported a test-retest effect size of −.13.

**Validity/Relationship to Real-World Outcomes.** Patterson et al. compared MMAA scores of older adults with schizophrenia or schizoaffective disorder with those of age-comparable normal controls. Results indicated that the schizophrenia group scored significantly lower than the normal controls. Additionally, statistically significant differences were found between the schizophrenia group and the normal controls on medication adherence.

The MMAA allows participants to read actual prescription labels, indicating that the assessment has good ecological validity. Construct validity for the MMAA appears to be good because scores on the MMAA were found to be significantly correlated with scores on 2 other performance-based measures of functionality, the direct assessment of functional status and the QWB. Construct validity was further examined by comparing performance on the Mini Mental State Examination (MMAA) with and a measure of cognitive impairment and negative symptoms of schizophrenia. This analysis further yielded significant correlations, in which lower scores on the MMAA were associated with increased cognitive impairment and negative symptoms. In the MATRICS study, when examining the relationship of the measures to cognitive performance, the MMAA performed better than the following alternative assessment measures: Maryland Assessment of Social Competence (performance-based measure), Schizophrenia Cognition Rating Scale (interview-based measure), Clinical Global Impression of Cognition in Schizophrenia (interview-based measure), and the Global Assessment of Functioning (GAF) (clinician-rated measure).

Unlike its sister measures (the UPSA and SSPA), the MMAA has relatively few studies demonstrating its relation to real-world outcomes. The MATRICS team found that the MMAA was modestly related to self-reported measures of global (r = −.18) and work functioning (r = −.26) but unrelated to self-reported social and independent living.

**Clinical Trials/Sensitivity to Change.** The MMAA was utilized in the FAST and PEDAL trials. Results of the FAST study indicated that there were no significant changes (P = .27) in MMAA performance between the behavioral intervention participants and the support-group intervention participants. Additionally, results of the PEDAL study found no significant differences between the PEDAL condition and the support condition after 6 and 12 months (P = .07 and P = .40, respectively).
However, significant differences were found between the 2 conditions at 18 months \( (P = .04) \), with the PEDAL condition outperforming the support condition.

Test of Adaptive Behavior in Schizophrenia

The Test of Adaptive Behavior in Schizophrenia (TABS)\(^\text{20}\) was developed in 2007 as a means of assessing functional capacity and 2 additional domains deemed important to functioning: \( (a) \) initiation of tasks in the real world and \( (b) \) identifying problems that occur during the course of functional activities. As such, the TABS consists of 5 subtests assessing the person’s ability in the following domains: \( (a) \) Medication Management Skills (ie, taking 3 types of medication as directed, identifying that the pills would run out prior to the end of the week, and generating a solution to this problem), \( (b) \) Empty Bathroom (ie, identify what items are needed in an empty bathroom so they could use it every day), \( (c) \) Shopping Skills (ie, using a map to get to a store, shop for items by viewing pictures of supermarket aisles, and identify when they have not received correct change), \( (d) \) Clothes Closet (ie, selecting appropriate clothes for various circumstances), \( (e) \) Work and Productivity (ie, correctly collate flyers), and \( (f) \) Social Skills (ie, tester rating of testee’s eye contact, speech, etc., during the test). Participants receive a “percent correct” score for each of the 6 domains, with the average of the 6 domain scores serving as a total score \( \text{range} = 0–100 \).

Reliability. All reliability data appear in the study reported by Velligan and colleagues.\(^\text{26}\) In this study, internal consistency for the TABS was good, with Cronbach alpha for all items reported as .84. When using the 6 subscale scores, alpha was reported as .80. Test-retest reliability over a 3-month period was also good, with the Pearson correlation coefficient and intraclass coefficient both being .80.

Validity/Relationship to Real-World Outcomes. TABS scores also appeared related to other measures of functional outcome, with significant correlations being found between the TABS and the Multnomah Community Ability Scale (.49), the Social and Occupational Functioning Scale (.43), and the Independent Living Scales (ILS; .58). As expected, TABS scores were significantly related to negative symptoms \(-.56\) but not to positive symptoms \(-.12\). When scores on the TABS were compared between persons with schizophrenia and normal controls, those with schizophrenia demonstrated significantly worse performance.

The TABS is a newly developed scale and, therefore, has not been tested for its association with real-world outcomes such as employment, residential independence, or education.

Sensitivity to Change/Use in Clinical Trials. To date, there have been no published clinical trials in which the TABS has been used as the primary outcome.

Independent Living Skills Survey (Self-Report Version)

The Independent Living Skills Survey (ILSS)\(^\text{21}\) is administered via 1 of 2 versions, an informant report form (ILSS-I) and a self-report form (ILSS-SR). We describe the self-report version here. The ILSS-SR consists of 70 items assessing 10 domains of functioning: \( (a) \) Personal Hygiene, \( (b) \) Appearance and Care of Clothing, \( (c) \) Care of Personal Possessions and Living Space, \( (d) \) Food Preparation, \( (e) \) Care of Personal Health and Safety, \( (f) \) Money Management, \( (g) \) Transportation, \( (h) \) Leisure and Recreational Activities, \( (i) \) Job Seeking, and \( (j) \) Job Maintenance. Response options are “yes” or “no” with regard to whether or not the individual had performed the skill during the past month. Scores for each domain are developed by averaging the domain items.

Ease of Administration. The ILSS-SR requires no specialized qualifications for administration and can be administered by a suitably trained paraprofessional. No props are required, and the test can be administered in approximately 20–30 min. Therefore, the ILSS-SR received a relatively high Ease of Administration grade.

Reliability. The ILSS-SR was initially validated on 3 samples of participants with “Severe and Persistent Mental Illness” \( \text{total} \ N = 448 \).\(^\text{21}\) Internally reliability for the 10 subscales was mixed, ranging from .435 for the Transportation subscale to .90 for the Job Seeking subscale. A total of 6 subscales had coefficients below .70. In a later study of middle-aged and older adults with psychosis, Perivoliotis et al\(^\text{22}\) found similar internal reliability coefficients, which ranged from .48 for the Transportation subscale to .90 for the Job Seeking subscale, with 6 subscales having reliability estimates below .70.

Test-retest reliability was conducted on a series of four 6-month intervals. Test-retest coefficients ranged from .418 for the Leisure and Recreational Activities subscale to .904 for the Job Maintenance subscale. As with the internal reliability coefficients, a number of scales had coefficients below .70 \( (n = 7) \).

Validity/Relationship to Real-World Outcomes. Validity for the ILSS-SR was evaluated by correlating total scores and scores for each functional area of the ILSS-SR with scores on the Brief Psychiatric Rating Scale (BPRS), the Profile of Adaptation to Life—Change version, and the Global Assessment Scale (GAS).\(^\text{21}\) The correlation between BPRS Total score and ILSS-SR Total score was -.318, indicating that greater symptoms of psychosis were associated with worse overall functioning. Similarly,
the GAS and ILSS-SR Total score were correlated at .375. Perivoliotis and colleagues\textsuperscript{22} compared ILSS-SR scores of 57 middle-aged and older adults with psychosis with a sample of 40 nonpsychiatric participants of similar age. The psychiatric sample scores were significantly worse than the nonpsychiatric sample on 8 of 10 subscales, and the 2 samples did not significantly differ on the Care of Personal Possessions and Job Maintenance subscales. Also, participants with psychosis had significantly lower ILSS-SR Total scores.

In the Perivoliotis sample, ILSS-SR scores were correlated with Positive and Negative Syndrome Scale (PANSS) scores. Results of these analyses indicated that only the Food Preparation and Storage subscale of the ILSS-SR had a significant negative correlation with PANSS positive symptoms.

Wallace and colleagues\textsuperscript{23} examined the relationship between ILSS-SR score and 2 measures of employment. The first was whether or not the participant worked at least 1 month and earned a minimum of $300 during the month of work. The second definition was whether or not the participant “worked at that income level for either 6 continuous months or the last 3 months of the follow-up period” (ie, sustained employability). Each of these definitions was entered as a dependent variable in separate stepwise survival analyses, with ILSS-SR Total score entered as the primary independent variable. Other independent variables included a 3- or 15-day sample of actual work behavior, the BPRS, the Social Adjustment Scale-II, a measure of work history, and initial eligibility status (adjudicated disabled or not). The authors reported that the ILSS-SR was the best predictor of employability and second best predictor of sustained employability. Although these data appear encouraging, no data were provided to indicate the significance of these findings or the magnitude of the effect.

Use in Clinical Trials/Sensitivity to Change. Liberman and colleagues\textsuperscript{23} compared the efficacy of a skills-training intervention with that of an occupational therapy intervention for producing change in ILSS-SR scores. Over the course of 6 months (3 h per day, 4 days per week), participants in the skills-training intervention were taught social and independent living skills such as basic conversation, recreation, medication management, and symptom management. The occupational therapy condition was equivalent in time, except participants engaged in expressive and artistic activities and encouraged to “individualize their interests and abilities through arts and crafts, discussion of feelings, and articulation of personal goals.” Results indicated that participants in the skills training condition showed significantly greater change in ILSS-SR Total scores over a 2-year follow-up period than those in the occupational therapy condition. For individual ILSS-SR subscales, participants in the skills-training condition showed significantly greater improvement on the Care of Personal Possessions, Food Preparation, and Money Management scales. The authors reported that change was most pronounced during the active phase of the intervention (baseline to 6 months), with a tapering effect occurring following treatment.

Wallace and colleagues\textsuperscript{21} replicated the design of the Liberman study in a sample of 131 community-dwelling individuals with schizophrenia. In this study, a third intervention condition, routine services, was added. Results indicated that over the course of 36 months, participants in the skills-training intervention scored significantly higher on the ILSS-SR Food Preparation, Money Management, and Health Maintenance scales than those in either the occupational therapy or routine services conditions.

In their CBSST trial (described above), Granholm and colleagues\textsuperscript{13} compared participants in the CBSST condition with those in the treatment as usual condition on change in ILSS-SR Total scores. In this study, a modified Total score was used which consisted of scores on 5 subscales: (a) Appearance and Clothing, (b) Personal Hygiene, (c) Health Maintenance, (d) Transportation, and (e) Leisure and Community. The remaining 5 scales were omitted because participants resided primarily in board and care settings and/or were retired. Results indicated that participants in the CBSST condition had significantly higher ILSS-SR scores at posttreatment than those in the treatment as usual condition.

Global Assessment of Functioning

The GAF scale is a clinician-rated measure of an individual’s overall functioning. The GAF was originally utilized in the Diagnostic and Statistical Manual of Mental Disorders, Third Edition Revised (DSM-III-R) and continues to be used as Axis V of the Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision). The GAF consists of 10 “ranges of functioning,” each 10 points, that describe a person’s overall psychological, social, and occupational functioning. Clinicians assign a value ranging from 1 to 100 with 100 indicating superior functioning.

Ease of Administration. GAF scores are typically assigned based upon a clinical interview and mental status examination, which vary in terms of administration time. However, published reports indicate that GAF scores can be ascertained in approximately 5–15 min.\textsuperscript{23} Because the GAF is not a “test” in the typical sense, it requires no special props to administer and modest training is required.\textsuperscript{24}

Reliability Data are scant on the reliability of the GAF for exclusively psychotic disorders. Startup et al\textsuperscript{25} examined the reliability of the GAF scale in 64 participants diagnosed with schizophrenia. Interrater reliability coefficients ranged from .89 to .95 over the course of this
1-year study. Other studies report on its reliability in general mental health populations. Endicott et al.\textsuperscript{26} examined the reliability of the GAS, a precursor to the GAF, and found interrater reliability coefficients ranging from .61 to .91 in 5 separate studies. Jones and colleagues\textsuperscript{27} found interrater reliability coefficients ranging from .56 to .76 in a sample of 102 hospital patients in England, of which 75 were diagnosed with schizophrenia. Hilsenroth et al.\textsuperscript{28} found the interrater reliability of the GAF to be .86–.92 in a sample of 44 individuals admitted to an outpatient university-based community. Diagnoses in this sample were primarily mood disorder ($N = 25$). A review article reports the interrater reliability of the GAF ranging from .62 to .96.\textsuperscript{24} In terms of practice effects, the MATRICS team\textsuperscript{5} reported an effect size of .04.

**Validity/Relation to Real-World Outcomes** Few studies describe the validity of the GAF for individuals diagnosed with schizophrenia or severe psychosis. Startup and colleagues\textsuperscript{25} found that GAF ratings were generally unrelated to concurrent measures of clinical symptoms or social behavior prior to receiving treatment. However, GAF scores were highly negatively correlated with clinical symptoms and social functioning following treatment. Two other studies compared GAF scores with other measures of functioning. Roy-Byrne and colleagues\textsuperscript{29} conducted a study of 337 psychiatric inpatients in which a revised version of the GAF was used. This revised version considered only social and occupational functioning, and not psychological symptoms, in rating overall functioning. In this study, GAF ratings provided by board certified psychiatrists did not significantly differ in those who were homeless ($N = 58$) vs those who had a residence ($N = 279$).

The second study\textsuperscript{30} focused on individuals with schizophrenia ($N = 53$) or schizoaffective disorder ($N = 109$) who had a minimum of 2-year history of continuous care. GAF scores were provided by clinical psychiatrists who were blind to the study design. GAF scores for 3 separate groups were compared, that is those requiring: (a) maximum supervision (ie, “24-h hospitalization or staff supervision”), (b) moderate supervision (ie, “daily contact with staff or significant others in a residential program or family care setting”), or (c) minimum supervision (ie, “intermittent contact with case managers, presence of roommates, or living alone in an unsupervised apartment”). Significant differences were observed between the minimum supervision group and both the moderate and maximum supervision groups. No significant GAF score difference was observed between the moderate and maximum supervision groups.

**Sensitivity to Change/Use in Clinical Trials** In their review of 10 treatment studies, Burlingame and colleagues\textsuperscript{24} found the GAF to be highly sensitive to change, with an average Cohen $d$ effect size of 1.10 across the studies.

**The ILS**

The ILS\textsuperscript{31} measures an individual’s competency in performing daily functioning activities and for caring for themselves. The ILS consists of 5 subscales: (a) Memory/Orientation, (b) Money Management, (c) Management of Home and Transportation, (d) Health and Safety, and (e) Social Adjustment. Scores on these 5 subscales are added to create an overall score ranging from 0 to 140, with higher scores indicating better performance. The ILS also consists of 2 factor-analyzed subscales: (a) Problem Solving and (b) Performance/Information. The Problem Solving subscale (ILS-PB) is commonly used as a measure of problem solving required of individuals independently residing in the community\textsuperscript{30,32} and consists of 33 items from the 5 domains tested by the ILS. As per the scoring manual,\textsuperscript{31} standard scores on the ILS-PB may be used to make decisions regarding level of supervision an individual requires in the community. Scores ranging from 20 to 39 indicate maximum supervision, scores from 40 to 49 indicate moderate supervision, and scores from 50 to 63 indicate minimum supervision.

**Ease of Administration.** According to the testing manual, the ILS requires approximately 45 min to administer and 10 min to score.\textsuperscript{31} However, Revheim and Medalia\textsuperscript{30} indicate that the ILS Problem Solving scale (ILS-PB) requires 20–25 min to administer. The testing kit contains props (eg, driver’s license, credit card, key) and is not in the public domain. The kit can be purchased from the publisher.

**Reliability.** Revheim and Medalia\textsuperscript{30} report solid reliability for the ILS. Specifically, the coefficient alpha for the entire ILS is .86, test-retest reliability is .90, and interrater reliability is .98.

**Validity/Relation to Real-World Outcomes.** Standard validity checks on the ILS indicate that clinical samples score significantly lower than nonclinical samples (discriminant validity).\textsuperscript{30,32} Revheim and Medalia\textsuperscript{30} also indicate that ILS scores were strongly correlated with Wechsler adult intelligence Scale-revised comprehension scores ($r = .65$). Revheim et al.\textsuperscript{33} report significant correlations between several neurocognitive tests and ILS-PB scores.

The ILS-PB has been associated with various real-world outcomes. In 2 separate studies, significant differences were found in ILS-PB scores for inpatients and outpatients with schizophrenia or schizoaffective disorder,\textsuperscript{32,33} suggesting that the ILS-PB was sensitive to detecting residential status. A third study compared ILS-PB scores for participants requiring 3 levels of supervision in their living situation, as determined by chart review. Levels of supervision were (a) maximum supervision (24-h hospitalization or staff supervision), (b) moderate supervision (daily contact with staff or significant others...
in a residential program or family care setting), and (c) minimum supervision (intermittent contact with case managers, presence of roommates, or living alone in an unsupervised apartment). Results indicated that those requiring maximum supervision scored significantly lower than those requiring moderate or minimum supervision. The moderate supervision group also scored significantly lower than the minimum supervision group. These results provide confirmation on the ability of the ILS-PB for detecting residential living situation.

**Sensitivity to Change/Use in Clinical Trials.** Medalia et al34 conducted a 3-group randomized controlled trial of 54 participants with schizophrenia or schizoaffective disorder, in which the aim was to remediate problem-solving skills. Group 1 (N = 18) was a problem-solving remediation group, in which participants worked with a software program called “Where in the USA is Carmen Sandiego?” to solve a simulated crime case. Group 2 (N = 18) was a memory group, which utilized a software package titled, “Memory Package” encompassing several tasks to teach memory skills and strategies for remembering. Group 3 was a control group (N = 18), in which participants engaged in routine unit activities such as arts and crafts. All 3 groups consisted of twice weekly sessions (25-min each) for 5 weeks. Participants completed the ILS-PB prior to and following the intervention. Results indicated that participants in the problem-solving remediation group showed significantly greater increases in ILS-PB scores than both the memory and control groups.

**Discussion**

This review explored the existing data for several measures intended to assess functional recovery in persons with psychosis. As described by Harvey and Bellack (this issue), an adequate definition of functional recovery can be elusive. Harvey and Bellack differentiated clinical remission from functional recovery by indicating that the latter is related to “the performance of daily activities that are required for self-maintenance (earning an income and maintaining a residence).” Clearly, other factors should be of interest to professionals working with this population (eg, symptoms of the illness and quality of well-being), but as Harvey and Bellack indicate, the extant literature has suggested that improvement/remission of the hallmark symptoms of psychotic illnesses (eg, hallucinations and delusions) is not necessarily linked to improved functioning nor does there appear to be close links between either of these factors and well-being. Therefore, recovery in those domains appears to require not only separate assessments but also different approaches than what would be used for improving functioning.

Measures were reviewed based upon their Ease of Administration, Reliability, Validity/Relationship to Real-World Outcomes, and Sensitivity to Change/Use in Clinical Trials. What is clear from the existing literature is that no “gold standard” measure has been developed. In particular, most research with these scales has been cross-sectional, and very little is known about their validity for predicting future real-world outcomes such as one’s prospective ability to reside independently or sustain employment. This clearly demonstrates that more research on these existing measures is needed, as is a greater emphasis on developing new measures for assessing functional recovery. Funding for both of these needs is very much needed as a means of encouraging quality research that ultimately benefits those in need. Toward this goal, the National Institute of Mental Health has funded the “Validation of Intermediate Measures” study, which will include 160 individuals with schizophrenia and will assess measures of functioning for test-retest reliability, relationships to community functioning and cognitive performance, and its practicality and tolerability. Indeed, the validation of intermediate measures study will include measures described in this study (eg. UPSA, UPSA-Brief, TABS) as well as new measures not included in this review.

What has been learned regarding the existing measures? First is that the UPSA has received the most attention in the research literature. A number of studies describe the UPSA’s validity/relationship to real-world outcomes (eg, residential independence and community responsibilities) and sensitivity to change in clinical trials. Given its superiority in terms of literature and existing evidence, the UPSA is, perhaps, the best available measure for determining functional recovery in people with psychosis. The UPSA’s primary disadvantage is that it requires several props that may make its administration difficult. As mentioned above, an additional limitation is that studies describing the UPSA’s relation to real-world outcomes have been cross-sectional rather than longitudinal, making it unclear if improvement on the UPSA would equate to an individual’s acquired ability to reside independently, for example. When time limitations exist, professionals might consider the UPSA-Brief, which is a shorter and more “portable” version of the original UPSA. This scale has clear advantages in terms of its shorter administration time (ie, 10–15 min) and fewer props. Like the UPSA, the UPSA-Brief also appears limited by its cross-sectional relationship to residential independence. In addition, other than one study demonstrating that the scale may be sensitive to change,15 no clinical trials have used the UPSA-Brief as a primary outcome measure. Therefore, it is unknown whether the correlations between measures would translate to sensitivity to change; this is particularly problematic if it is employed as an outcome measure for an intervention that targets functional domains trimmed from the full version.

The remaining scales appear to lack the overall scope of research devoted to the UPSA and UPSA-Brief. While some scales appear advantageous in terms of Ease of
Administration and Reliability, they lack enough solid evidence that they are related to real-world outcomes or sensitive to change (eg, SSPA). The TABS, which appears similar to the UPSA in terms of its Ease of Administration and Reliability, has not yet been tested in terms of its relation to real-world outcomes or sensitivity to change via clinical intervention. By no means, do we discourage use of these scales. Rather, we strongly encourage researchers to systematically examine the psychometric properties of these scales to more clearly understand their place in the assessment of functional recovery.

The following are additional points to consider with regard to assessing functional recovery. First, as discussed by Harvey and Bellack, 2 aspects of recovery exist. One is the person’s increased capacity to engage in functional behaviors and another is the person’s desire or will to do so. While many of the measures described here appear to adequately assess the prior, there is a lack of attention to the latter. People who have the capacity to reside independently but not the will to do so would only have achieved half of what is necessary for “functional recovery” to occur. The TABS, which is a performance-based measure of functional capacity, has incorporated tasks to assess the individual’s motivation, but more research is needed and greater development of assessing this aspect of recovery is needed.

Second, Harvey and Bellack discuss the important issue of a “time frame” to index functional recovery. Achieving capacity to reside independently, eg, not only takes greater time and commitment to achieve than remission of clinical symptoms but also becomes more difficult to assess via the measures described in this review. Indeed, none of these measures adequately assesses long-term functional recovery, instead relying on a momentary snapshot of the individual’s current functioning. Further research is needed to understand how these measures can be used to assess ongoing recovery and sustained recovery. How might this be done in the context of research examination? Should individuals be repeatedly tested every 2 weeks? Every month? How do the current scales respond to repeated administration and do people learn how to complete the test but not master the skill? These are important questions for our field that are in need of answers.

Third, the scales described in this review are tailored for assessing functioning in Western cultures. To be sure, even in Western cultures, the number of desirable functional outcomes (eg, homemaking, outside employment, residential independence) goes beyond the skills directly assessed by these scales, and in many cases, the definition of functional recovery may take different forms depending on individual and cultural factors. Perhaps, an excellent example of these 2 points is the studies by Cardenas and Mausbach. In the prior, Cardenas argues that in Latino cultures, cultural factors may dictate that attainment of employment, but not necessarily residential independence, may be an ultimate goal. In the study by Mausbach, the culture of some residential facilities (eg, board and cares) combined with motivational factors for achieving functional outcomes (eg, residential independence or employment) may lead to increased capacity for these outcomes, but not necessarily result in ability to actually achieve this greater autonomy.

In sum, this review appears to raise as many questions as those it answered. What began as an intent to aid professionals and researchers to select a functional recovery measure ended up finding a need for additional research on these and new measures of functional recovery. Ultimately, no gold standard measure exists, but of those available, we find that the most utilized and valid measure of functioning is the UPSA. Indeed, a relatively great deal of literature supports its usefulness for assessing functional recovery. Our hope is that all the measures described in this review will receive greater attention in the scientific literature and that more funding be available to develop these and new measures of functioning.

Funding
National Institute of Mental Health (MH078775 to T.L.P.).

Acknowledgments
Conflict of Interest: Dr Patterson developed 4 tests reviewed in this article (ie, the UPSA, UPSA-Brief, SSPA, and MMAA) and Dr Mausbach has published several articles assessing these tests. Neither Dr Patterson nor Dr Mausbach has received financial compensation for the development or administration of these tests.

References
6. Twaemley EW, Doshi RR, Nayak GV, et al. Generalized cognitive impairments, ability to perform everyday tasks, and...