Performance-Based Measures of Functional Skills: Usefulness in Clinical Treatment Studies

Philip D. Harvey1–3, Dawn I. Velligan4, and Alan S. Bellack5,6

2Department of Psychiatry, Mt Sinai School of Medicine; 3Department of Veterans Affairs VISN-3 Mental Illness Research, Education, and Clinical Center; 4University of Texas Health Science Center at San Antonio; 5Department of Veterans Affairs VISN-5 Mental Illness Research, Education, and Clinical Center; 6Department of Psychiatry, University of Maryland School of Medicine

Recently, attention to the assessment and treatment of functional disability has increased notably. It is widely understood that impairments in everyday living skills, including independent living skills, social functions, vocational functioning, and self-care, are present in people with schizophrenia. It has also become clear recently that assessment of these skills can pose substantial challenges. These challenges include selection of meaningful short-term outcome measures and avoiding bias and reduced validity in the data. Self-report, direct observation, and informant reports of everyday disability all have certain advantages but appear to be inferior to direct assessment of skills with performance-based measures. This review outlines the issues associated with the assessment of functional skills and everyday functioning and provides a description of the strengths and weaknesses of these approaches. We conclude that direct assessment of functional capacity has substantial advantages over other measures and may actually provide a more direct and valid estimate of functional disability than performance on the more distal neuropsychological assessment measures.

Key words: schizophrenia/disability/outcomes/neuropsychology

Functional Disability in Schizophrenia

There has been an increased sensitization on the part of the research and clinical communities to the fact that schizophrenia is more than an illness of delusions and hallucinations. Further, the realization is now widespread that the treatment of delusions and hallucinations does not guarantee more global improvements in functioning and subjective quality of life. As a result, improving everyday functioning and subjective appraisal of everyday functioning have been suggested as a viable target for pharmacological, behavioral, and rehabilitation-focused interventions. In order for treatments to be thoroughly evaluated in terms of their short-term efficacy and longer term effectiveness, it is necessary to have outcome measures that validly capture both the domains of everyday disability and how they respond to treatment interventions. As we will indicate below, different measures may be required to capture these 2 aspects of everyday functional disability.

Another domain of impairment in schizophrenia, which falls outside the objective indices of disability that we review, is subjective quality of life. This intrinsically self-reported feature of the illness has been studied extensively. Subjective quality of life may impact on a number of features of schizophrenia, including adherence to medication and other aspects of treatment interactions. We will not be discussing these issues because of our space considerations.

In tribute to Wayne Fenton, he was clearly aware of and in agreement with the statements in the first paragraph. Wayne was at the forefront of the movement to assess and treat functional disability. Through his support of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) and Treatment units for Remediation of Neuropsychological Impairment in Schizophrenia (TURNS) initiatives and his convening work groups at the National Institutes of Health to write a white paper on functional disability in schizophrenia, he put substantial momentum into the movement toward direct assessment and treatment of functional disability in schizophrenia. Supporting both pharmacological and behavioral interventions with disability as the ultimate treatment target, he was acutely aware of the need to change outcomes in schizophrenia. Patients with schizophrenia will be better off in the long term because of Wayne, and his loss, tragic as it is, came after his initiatives in this area had taken shape.

Multiple domains of everyday functional outcome are impaired in schizophrenia. Rates of competitive employment

1To whom correspondence should be addressed; tel: +1-212-659-8713, fax: +1-212-996-8931, e-mail: Philipdharvey1@cs.com.
are under 20%, and even in supported employment, job tenure lasts only a few months on average with as many as 50% of patients who attain work having unsatisfactory job terminations.1–3 Patients achieve lower educational levels than would be expected by socioeconomic status and other demographic variables.4 A variety of deficits are observed in independent living skills including the abilities related to using public transportation, cooking, care of living quarters, money management, and medication adherence.5–8 Some more chronic patients even manifest substantial impairments in activities of daily living,9 and the major medical morbidity associated with schizophrenia is at least partly due to reduced tendencies on the part of schizophrenia patients to spontaneously seek health care, particularly preventative care. Further, the level of poverty experienced by many individuals with schizophrenia because of other components of their disability leads to reductions in the availability and the quality of health care received.

As many as two-thirds of schizophrenia patients are unable to fulfill basic social roles, such as spouse, parent, and worker, even when psychotic symptoms are in remission. Most patients have significant impairments in social relationships. They often are isolated and when they do interact with others, they have difficulty maintaining appropriate conversations, expressing their needs and feelings, achieving social goals, or developing close relationships. Social networks of patients with schizophrenia are also smaller than those of control subjects.4,10 Schizophrenia patients are not grossly odd or impaired as children, but subtle signs of interpersonal difficulty are often present in childhood or adolescence.11 These problems tend to become more apparent during the early stages of the disorder, when many patients become progressively more isolated and have increasing difficulty maintaining previous levels of social adjustment. Premorbid social competence is among the best predictors of long-term outcome, either because poor premorbid adjustment is a marker of a more pernicious form of illness or because social competence is a coping skill that helps the individual achieve goals and avoid stress.

These multiple domains of impairment combine to make schizophrenia among the top disabling conditions worldwide for young adults, according to the World Health Organization.12–13 While psychiatry has been successful in treating the positive symptoms of schizophrenia, not much progress has been made in treating the major role impairments associated with the disorder.

In summary, there are several domains of disability in schizophrenia, and the prevalence of disability is quite high. Compared with the prevalence of treatment-resistant positive symptoms, which are estimated at about 35% of patients with the illness, these functional deficits are apparently more common. Functional impairments also contribute to the indirect cost of the illness, which by most estimates, is about double the direct cost of illness. In combination, these costs have been estimated to be as high as $60 billion annually in the United States.14–16 Thus, there is no question that the treatment of functional disability is an urgent priority.

Determinants of Functional Disability

There are many different reasons that an individual may be unemployed, homeless, and lacking friends and other social contacts. Living in an impoverished area or being a recent immigrant and lacking appropriate language skills can easily result in poor vocational, residential, or social outcomes, as could prejudice and lack of educational opportunities. However, even people with schizophrenia who have access to opportunities and resources typically have substandard achievement (given expectations based on their parents’ achievements). There are clearly features of the illness that are correlated with tendencies toward functional disability. Cognitive deficits in several different ability areas are believed to underlie much of the significant functional impairments observed in schizophrenia.5,17 Deficits in psychomotor speed, attention, memory, and executive functions have been found to predict community outcome, social skills deficits, ability to learn in rehabilitation programs, and quality of work. In fact, cognitive deficits are the single best predictor, among the many features of schizophrenia, of functional disability.18 At the same time, some other symptoms, such as psychosis, can lead to problems in sustaining functional goals through interpersonal conflicts that can lead to job loss, eviction, and social rejection because of aversive interactions.

There are substantial efforts currently underway to reduce disability through the treatment of cognitive deficits. These initiatives are reviewed elsewhere in this issue. Regardless of intervention strategy, we believe that direct assessment of disability as a treatment outcome is important. First, the amount of variance in real-world functional disability accounted for by cognitive impairments is variable, with some abilities sharing as little as 10% of variance with everyday outcomes.19 Second, there is little evidence, particularly pharmacological evidence, that improving cognition would change functional outcomes for the better. The frequently reported cross-sectional correlational relationships between cognition and disability may be due to other, as yet unidentified, factors. Finally, mediating factors may influence everyday outcomes in ways that are as yet unknown. It is also possible that improving cognitive functioning would have an adverse impact on some mediating factors, possibly by increasing awareness of disability and leading to an increase in subjective dissatisfaction.

The Competence/Performance Distinction

For the most part, efforts to assess disability directly have focused on the skills or abilities that are assumed
to underlie what people do. Critical to any attempt to accurately measure disability with behavioral assessments of skills is the separation of assessment domains into what the person can do and what they actually do. This is referred to as the “competence/performance distinction” and is critical for comprehensive understanding of functional disability.20 There are many influences in real-world outcome other than basic personal competencies (or capacities), and functional capacity is not likely to completely overlap with real-world performance. This lack of overlap may be due to both characteristics of the individuals’ themselves as well as a host of environmental factors.

Among the personal characteristics that may influence real-world performance above and beyond the influence of functional capacity are confidence, motivation, willingness to take risks, and the ability to self-evaluate and self-monitor. Self-evaluation, often referred to as “meta-cognition,” has been shown to be particularly impaired in people with schizophrenia,21,22 although even healthy people demonstrate consistent deficits in accurate self-evaluation.23 Impairments in monitoring can lead to bidirectional misestimation of skill levels, in that overestimation of skills can lead to attempting excessively challenging tasks and underestimation can lead to unwillingness to make efforts in situations where success may occur.

Some studies have found that social and environmental factors such as disability status or racial characteristics are correlated with poor work outcomes.24 In contrast to individual features such as confidence, motivation, or self-monitoring, these are not factors which can be treated in individual patients with direct interventions. Thus, for the purposes of clinical treatment studies of interventions attempting to reduce disability, these factors are not likely to be part of the treatment intervention, although they are likely to be influences on the overall picture of functional disability.

Thus, we are suggesting that researchers consider how to best measure disability and reduction of disability in a direct manner. There are several means to do this, and we argue that performance-based measures of functional capacity, in domains of social, vocational, and everyday living skills, may be suitable for use in both research and clinical settings. In making this point, we review the other methods for assessing disability, with an eye toward their utility in clinical trials, and evaluate the validity of performance-based measures. We will consider a number of features of these assessments that are directly relevant to their use in treatment studies.

Methods for Assessment of Functional Disability

The assessment of functional outcome in schizophrenia is complicated. Multiple instruments are available for assessing a variety of outcome domains, including both indices of competence as well as real-world outcomes. It is important to separate the content domains of outcome and potential disability, which include social, vocational, self-care, and independent living, from the assessment methods employed. These assessment methods include global rating scales that are designed to be rated on the basis of all available information, self-report instruments, direct observation of behavior, informant reports, and performance-based measures of functional skills.

Each of these methods has strengths and weaknesses. Self-report would be desirable because there are many behaviors to which only the person being assessed has access. Direct observation has the benefit of the ability to develop highly objective and reliable coding systems. Informant reports have the benefit of avoiding many response biases in other methods, and contrasts across different informants may provide a broad-brush picture on variance in performance across different settings. Performance-based measures eliminate the possibility of response biases and can be adjusted to capture highly specific features of functional skills. In regard to limitations, self-report in schizophrenia patients is particularly biased, for several reasons. Direct observation is not useful for many behaviors, and for patients who are socially isolated and unemployed, the frequency of the behaviors to be observed is quite low. Many people with schizophrenia lack key informants, and the amount of contact informants have with the patient can be quite variable. Performance-based measures of capacity require evidence of concurrent validity, and as noted above, there are other influences on everyday performance that can reduce the relationships between competence and performance to a substantial degree. Finally, as performance-based measures, there is a demand for the subjects to actively participate in the assessment process, and poor motivation or uncooperativeness could lead to low scores.

The Limitations of Self-Report of Functional Disability in Schizophrenia

It has long been known that people with schizophrenia manifest reduced awareness of the presence and significance of their psychotic symptoms.25 In fact, “lack of insight” may be the most common symptom of schizophrenia other than functional disability. Despite this finding, many functional outcome rating scales are designed to collect self-report information from people with schizophrenia about their functioning, including self-evaluation of social functioning and complex features of occupational and everyday living skills. Several recent studies have suggested that this may be a problematic way to assess functioning. For instance, McKibbin et al26 found that the correlation between self-reported illness burden and self-reported disability were quite substantial but that neither of these self-reports correlated with either neuropsychological (NP) performance or scores on a performance-based measure of functional
capacity. Consistent with this finding, Keefe et al\textsuperscript{27} found that correlation between self-reported cognitive deficits and performance on a NP assessment was $r = .04$ in a sample of schizophrenia patients. In contrast, an informant rating of cognitive impairment, provided by someone who was unaware of the patient’s NP performance, was correlated at $r = .40$ with NP performance. Interestingly, in that study, the single largest correlate of NP performance was scores on a performance-based measure of functional capacity. Finally, in a very recent study,\textsuperscript{28} schizophrenia patients and their case managers were asked to complete the same everyday functioning rating scale, while the patients also performed an NP and functional capacity assessment. For every domain of everyday functioning examined, the case manager’s reports and the patient’s reports were poorly correlated and only in the area of personal care (activities of daily living) were the correlations even statistically significant. When correlations between case manager and self-reports, NP performance, and functional capacity were calculated, the case manager report was more strongly correlated with the external validators for every functional domain.

These data suggest that, relative to the reports of specific informants, self-reports on the part of people with schizophrenia are quite problematic. On the surface, these findings would also suggest that requiring reports from informed observers, such as psychiatric case managers or caregiver relatives, would be a more appropriate strategy. Caregiver-rated scales have advantages over self-report measures in that the insight, memory, and assessment of caregivers may be more accurate than that of the patient. Moreover, these assessments, if done by family members, allow the rating of behaviors that treatment providers rarely have opportunity to observe. However, caregiver-rated scales and information provided by caregivers can be problematic in that different observers have different behavioral samples on which to base their responses, spend differing amounts of time with subjects, and may have different standards for appropriate performance. Caregivers may also not be sensitive to change. These differences among caregivers can create considerable variability in measurement. Moreover, as pointed out previously,\textsuperscript{29} as many as half to one-third of patients are not able to name a person who can supply this type of information.

One exception to this lack of available informants may be the case of patients who live in very supervised residences, such as locked board and care homes, nursing homes, or long-stay psychiatric hospitals. Several studies have reported high levels of interrater reliability and convergent validity with these rating, on specially constructed rating scales aimed at institutionalized patients.\textsuperscript{30} However, even those ratings can be flawed, and it cannot be assumed that because the staff knows the patient well they will generate valid ratings. Bowie et al\textsuperscript{31} studying convergence between clinical staff ratings and clinical researcher ratings of the same patients in nursing home care, found that these 2 ratings were very poorly correlated. Because the clinical researcher’s ratings of disability were more strongly correlated with NP performance than the clinical staff ratings, it can be assumed that the researcher ratings had greater validity.

Special Considerations in Treatment Studies

While the assessment of functional status in schizophrenia in general is challenging, there are additional demands imposed by treatment studies. Treatment studies require feasible and repeated assessments that can be reliably completed with fidelity to the protocol on specific visit dates. The content of the functional outcomes assessed has to be linked to the time frame of the trial and must examine aspects of performance that could realistically change due to treatment during the time frame of the study. For instance, marriage or establishment of some equivalent long-term relationship is not a reasonable outcome in a short-term treatment study nor is establishment of a permanent residence or obtaining full-time employment. Further, one of the other major hurdles in a short-term treatment study is identification of potentially beneficial effects of treatment in the context of environmental/social factors that may inhibit the expression of these gains. For instance, gains in employment-relevant skills may not be expressed in the real world, if the recipient of the gains is receiving full disability compensation, which would reduce the likelihood of attempts at employment. Another important factor that may confound interpretation of results from a treatment trial is uncontrolled concurrent treatment. For example, patients who are involved in a comprehensive vocational rehabilitation program\textsuperscript{32} or who begin drug treatment\textsuperscript{33} during a trial of a new antipsychotic may show gains that are greater than those who do not receive these ancillary treatments.

Another issue in treatment studies is the nature of the relationship between the individuals receiving treatment and the personnel at the treating site. While at many research sites, patients are involved in active treatment and are well known to the investigators, at other sites they may be recruited through advertisements. In such cases, even rating scales that suggest that all sources of information be used to generate ratings will be based largely if not entirely on the report of the patient. These ratings will, consequently, be vulnerable to all the concerns noted above, despite being characterized as “clinician ratings” and not self-reports of functional status.

These considerations suggest that direct measures of functional capacity, ie, what the patient can do, rather than indirect measures such as informant ratings or self-report, may be the most suitable candidates for outcome measures in clinical trials. These measures are less likely to be biased or to require unavailable informants, and they may be more sensitive to change. Despite these
advantages, there are still many considerations before these measures can be recommended without hesitation. These include feasibility, suitability for multisite studies, repeatability, sensitivity to change, and content validity. Each of these issues will be briefly described and then several examples of performance-based measures that could be considered for clinical trials are provided.

Feasibility. Clinical treatment studies have different timing demands than other types of clinical research. While many studies examining NP performance and functioning have used extensive assessments, the full assessment in the studies is often not completed in a single day. In contrast, baseline assessments for clinical treatment studies must be completed before the initiation of treatment and are often scheduled for a single visit. Because the functional outcome assessments are very unlikely to be the only evaluation performed, they must be able to be completed in conjunction with several other assessments. Thus, time to completion must be factored into the overall assessment burden at each assessment.

An additional factor is tester credentials and training. Many clinical trials sites will not have testers who are experienced in functional assessments and the tester/raters that they have may not have advanced degrees. If the assessment procedures cannot be learned and reliably administered by the typical research sites, the assessments may not be particularly practical.

Suitability for Multisite Studies. Because many treatment studies will be performed across multiple sites, instrument must be portable across the sites. This portability is one of the appealing features of NP assessments, which can almost by definition be easily employed across sites and raters. Some performance-based measures of functional skills have region-specific task demands, such as using specific maps or schedules. Similarly, because public transportation varies regionally, it is not an adaptive requirement in some regions of the country to be able to use public transportation whereas in others not being able to use such services means being essentially homebound. Social competencies may also vary regionally, and they certainly vary cross-culturally, so that studies that involve regions of the country with higher number of unassimilated immigrants may be different from regions where most people are locally reared. The complexities of multisite studies are increased markedly when the studies are performed multinationally, where both language and cultural norms are likely to impact on performance.

Repeatability. It is well known that cognitive assessments performed with NP tests do not yield the same scores across sequential assessments although the scores are quite highly correlated. Factors such as practice effects, regression to the mean, exposure effects, and random retest variation all operate to lead to differences in performance at reassessment above and beyond “true” differences. Performance-based measures of functional capacity would be subject to the same concerns. One solution to manage some of the retesting effects with NP measures includes alternate forms, which have their own issues, including nonequivalence. Performance-based measures rarely have alternative forms, so retesting effects have the potential to be a major concern, particularly with designs that are not simple parallel-group designs. At the same time, parallel research designs allow for the consideration of practice effects in the inactively treated group. Alternative forms are difficult to develop and may cause some measurement problems that are not addressable by between-group comparisons.

Sensitivity to Change. The central goal of a treatment study is to induce change in functioning, and outcome measures must be sensitive to change. Thus, the content of the assessment must be in domains that are amenable to change with interventions and not to those that are strongly linked to more dispositional factors. Dispositional factors such as intelligence will have high levels of temporal stability and test-retest reliability but are not likely to change in response to most treatment interventions. Skills to be examined in performance-based assessment should probably be selected for having only moderate correlations with likely immutable factors such as crystallized intelligence.

Content Validity. While it seems obvious that the competencies assessed in performance-based functional outcome measures should directly relate to everyday disability, the difficulties described above in defining the parameters of disability in schizophrenia make the evaluation of validity more challenging. What is clear, however, is that performance-based outcome measures need to measure competencies that are both required for every day functional success and impaired in schizophrenia. An additional issue is that the structure of performance-based tests of functional capacity is similar to the structure of NP tests. It is important that performance-based tests not just be more ecologically valid NP tests. Performance-based tests need to demonstrate that they provide information above and beyond that provided by NP tests with respect to real-world functional outcomes. For example, in a recent study, role-played measures of social competence contributed variance to the prediction of vocational functioning in schizophrenia above and beyond that associated with NP performance.

Review and Evaluation of Performance-Based Measures
Recently Moore et al. reviewed 31 different performance-based measures aimed at functional disability. Of these instruments, 20 were aimed at multiple functional
domains and the other 11 were single-domain instruments. The great majority of these instruments were aimed at dementia and normal aging, and few were actually developed for use in psychiatric conditions. Of the instruments that they reviewed, most were quite reliable, but the content was often aimed at self-care (including medical self-management) and not broadly aimed at the domains of functioning seen to be impaired in schizophrenia.

Rather than exhaustively review the available assessment instruments for nonpsychiatric populations, we will focus our comments on a few representative social, independent living, and vocational functioning instruments that have been employed in previous in studies of people with schizophrenia and that appear to be amenable for use as clinical outcome measures.

**Measures of Everyday Living Skills**

An important performance-based functional capacity measure was developed by Patterson et al. This scale, referred to as the University of California San Diego (UCSD) Performance-Based Skills Assessment (UPSA) was designed to address limitations in other performance-based assessments that had been validated with dementia patients. The scale takes about 30 minutes to complete and assess performance in 5 domains—household chores, communication, finance, transportation, and planning recreational activities.

Items in this scale involve performing a variety of skilled tasks, including manipulating money, making routine and emergency calls, reading maps and schedules, and performing shopping tasks. The scale is scored on a 0- to 100-point basis, with scores in the 5 domains set to 20 points each. Interrater reliabilities of ratings were excellent (intraclass correlation coefficient = .91) and test-retest reliabilities over a 1-week period of independent ratings were good (intraclass correlation coefficients ranged from .74 to .94), with the total score having a test-retest reliability of .93 at a 2-week retest.

A recent construct validity study suggested that scores on the UPSA (expressed as a total score) completely mediated the effects of NP impairment on a measure of everyday outcome rated by case managers. Interestingly, the zero-order correlations between NP performance and both UPSA scores and ratings of everyday outcome were substantial, while the correlations between negative symptoms, positive symptoms, and depression and UPSA scores were quite small. Thus, in that study, it appeared as if the UPSA had substantial construct validity as a measure of the functional outcome construct. Further evidence for the validity of the UPSA was provided in a recent study examining the sensitivity of UPSA scores to independence in residential situation. In that study, 434 participants were evaluated with the UPSA and 99 (23%) were living independently at the time of assessment. The discriminant validity of the UPSA was adequate (receiver operating characteristic [ROC] area under the curve = 0.74; 95% confidence interval: 0.68–0.79), with greatest dichotomization for the UPSA at a cutoff score of 75 (68% accuracy, 69% sensitivity, 66% specificity) or 80 (68% accuracy, 59% sensitivity, 76% specificity). The UPSA was also a significantly better predictor of living status than cognitive performance assessed by the Dementia Rating Scale (DRS; z = 2.43, P < .015).

In addition, a brief version of the UPSA, referred to as the UPSA-Brief has been developed. Using the same sample of 434 patients, a factor analysis of all the items was performed, finding a single-factor solution. The UPSA-Brief was created from the 2 subscales that loaded most heavily on this factor, namely the finance (factor loading = .85) and communication (factor loading = .80) subscales. Interestingly, these are the 2 shortest subscales on the instrument and require the fewest special props. The brief version of the UPSA was correlated quite strongly with total scores on the instrument, Pearson r = .91, as well as with concurrent validators such as the DRS (Pearson r = .59, P < .001). Finally, a cutoff score of 60 on the UPSA-Brief yielded accuracy of 70%, sensitivity of .82, and specificity of .58. An ROC analysis suggested that there was essentially no difference in the long and brief versions for their relationship with residential independence.

To date, the UPSA has not been used as a treatment outcome measure in a completed study. It is currently being utilized in treatment studies associated with the National Institute of Mental Health’s MATRICS-TURNS initiative. Information will be available soon as to the relative sensitivity of the UPSA (as well as the UPSA-Brief), compared with the MATRICS NP battery, to treatment-related changes induced by a pharmacological intervention.

An additional potential limitation of the current versions of the UPSA is a vulnerability to ceiling effects. Because the rest was developed for use in older patients, there is some potential for ceiling effects in younger patients. A potential solution to this problem is to include some supplemental, more challenging, items from other functional outcome rating scales.

The Test of Adaptive Behavior in Schizophrenia (TABS) is a performance-based measure of adaptive functioning designed to address limitations of other available measures including inadequate assessment of the ability to initiate and of the ability to identify problems that occur in the course of performing functional activities. For example, performance-based tests tend to ask a participant how they would solve a particular problem. However, it is not clear that the individual would have been able to identify the fact that a problem existed. In addition, because performance-based tests tend to ask the individual to respond to a contrived situation, there is little room to assess the individual’s ability to initiate or to adopt a flexible problem-solving style if necessary.
The TABS is designed to assess the abilities necessary to perform goal-directed activity including initiation, planning, problem identification, problem solving, sequencing, appropriate inhibition, and persistence in the context of 6 functional areas (work and productivity, medication management, independent living, shopping, basic hygiene, and social skills). TABS items demand considerable initiation (eg, spontaneously generating items that would be necessary to stock an empty bathroom), allow the subject the chance to identify specific problems on their own, prior to they being pointed out by the examiner (eg, identify that he or she was short changed, identifying that he or she will run out of medication), and provide additional points for spontaneously offering solutions (eg, spontaneously announcing a plan to remedy a problem with running out of medication). TABS scores are calculated as percentages with higher scores indicating better adaptive behavior.

In a study of 264 individuals with schizophrenia or schizoaffective disorders, tested on 2 occasions 3 months apart, the TABS demonstrated good test-retest reliability (.80) and interitem consistency (.84). Moreover, TABS scores were found to show good evidence of convergent validity, by being moderately to strongly correlated with other measures of functional outcome, negative symptoms, and neuropsychological test scores. However, measures of positive symptoms were not found to be related to TABS performance (discriminant validity). Thus, these preliminary data suggest that the TABS has reasonable evidence of construct validity. The sensitivity of the TABS to the effects of psychosocial or pharmacological treatments needs to be examined.

As reported in Velligan et al, one criticism that could be leveled at the TABS is that once a problem has been identified during the course of the first administration of the test (eg, not enough medication to fill the containers), the subject would be aware of the problem on repeated administrations. Repeated measurements of the TABS were obtained in the validation study at 3-month intervals. During that interval, there was little clinical evidence that subjects “remembered” problems that they had identified or that had been pointed out to them at previous testing sessions. Practice effects may be problematic in clinical trials with shorter intervals.

**Measures of Social Competence**

Behavioral observation strategies circumvent the limitations of self- and other reports (including interviews) by employing trained raters to evaluate what patients actually say and do based on direct observation of social interactions. Given the impracticality of conducting observations in the natural environment, these ratings characteristically are based on simulated interactions, or role plays. Role-play tests (RPTs) are among the most widely used behavioral assessment approaches. They have been used to assess many aspects of social functioning and treatment outcomes in the schizophrenia literature over the past 25 years. They have many advantages, including high degree of flexibility for use in different types of investigations, methodological rigor (in both administration and coding), and ability to assess social behaviors that are not readily available to other assessment strategies (eg, nonverbal and paralinguistic behaviors that are important components of social competence). Overall, the literature suggests that RPTs can be administered and coded reliably at different stages of illness, that they are sensitive to treatment effects, that they can differentiate diagnostic groups, and that they provide meaningful information about patients' response capability.

The Maryland Assessment of Social Competence (MASC) is the current iteration of a RPT originally developed empirically as part of a broader battery to assess social problem solving: the Social Problem Solving Assessment Battery. The battery was specifically designed for use with chronic psychiatric populations. It has high content and social validity and good test-retest reliability across different phases of illness. The role-play component involves administration of 4–6 three-minute simulated conversations with a confederate who is trained to facilitate the conversation in a standardized manner. It has been shown to detect differences between patients with schizophrenia and those with bipolar disorder and nonpatient controls. The RPT items in the original battery were selected to be broadly representative of conversation initiation and assertion situations experienced by people with schizophrenia. The MASC employs item content that is applicable for other specific social interactions. It has been employed recently in 5 large schizophrenia studies: a study of victimization in women who abuse drugs, a study of health care among people with diabetes, a study of vocational outcomes, a study of social skill among drug abusers, and a clinical trial comparing 2 antipsychotic medications. It was also examined as a proxy measure in the recent MATRICS project, and it is being employed as an outcome measure in the PURS project as well.

The MASC interactions are videotaped for subsequent coding by trained raters. Coding is customarily done by individuals who have not been involved in administration of the task, but that is not essential. There is a clear benefit in having different individuals administering and scoring this instrument in treatment trials because if the administrator does not know the criteria for scoring, their interaction with the patient would be less likely to induce better performance.

Coding can ordinarily be completed in little more than the time it takes to watch the videotapes, and thus, time required varies according to the number of scenes administered. Trained coders have been shown to regularly generate ratings that have high interrater reliability...
(eg, intraclass correlation coefficients above .70) for most behaviors and most scenes.33
The Social Skills Performance Assessment (SSPA)45 is an abbreviation and adaptation of the role-play components of the MASC. It uses only 2 of the interactions in the MASC, a social interaction (meeting a new neighbor) and an instrumental interaction (requesting attention from a landlord about a previously reported problem). As a result, the entire assessment can be performed in about 6 minutes, and the scoring is similar to the MASC in terms of the time required to score each of the vignettes. It can also be administered by raters who are not trained as scorers, which can be quite convenient for reducing potential biases in the administration of the test. The SSPA was successfully used as an outcome measure in a randomized clinical trial46 where it was found that scores on this measure were quite sensitive to the effects of treatment with atypical antipsychotic medications and actually were more sensitive to treatment than performance on an abbreviated battery of NP tests. Moreover, improvements on the NP tests were significantly correlated with improvements on the SSPA.

Vocational Skills
A final area of real-world functioning that could be assessed with performance-based measures in treatment studies is vocational skills and potential. The use of performance-based measures of vocational potential in treatment studies is important for 2 reasons. The first is the array of potential environmental influences on employment and the second is the time-course for real-world changes in employment status. It has already been shown that work functioning may be determined in large part by environmental and social factors, in that disability status is a potent predictor of the likelihood of employment.24 On the other hand, for patients with schizophrenia who are seeking work, there is evidence that cognitive remediation interventions, combined with intensive supportive employment, leads to substantial real-world employment gains.32,47 In terms of time-course, short-term treatment studies would be unable to detect several important employment-related outcomes. Both obtaining and sustaining employment are important outcomes, but evaluations of sustained success in employment are not feasible in short-term treatment studies.

Performance-based measures of occupational potential are actually in common use in vocational rehabilitation settings. The measure perhaps most amenable to use in treatment studies is the COMPASS system.48 The US Department of Labor lists 12 855 jobs in the “Dictionary of Occupational Titles” (DOT).49 Each job is assigned a DOT “Worker Qualifications Profile” in which it is rated with regard to the specific levels of all abilities that are required to perform it. These ability requirements are grouped under 3 general categories:

1. General Educational Development is comprised of Reasoning, Mathematics, and Language.
2. Aptitudes are the more specific abilities required by the job. They include Learning, Verbal, Numerical, Spatial, Form Perception, Clerical Perception, Motor Coordination, Finger Dexterity, Manual Dexterity, Eye-Hand-Foot Coordination, and Color Discrimination.
3. Worker Functions are requirements of the worker to perform in relationship to Data, People, and Things.

There are 6 functional levels for each of the general educational development abilities, 5 functional levels for the specific aptitudes, and 7–9 levels for worker functions. The Department of Labor has also determined the number of jobs within the US economy that require each of the various levels for each of the ability areas. In general, in the middle functional levels (where most people perform), thousands of jobs are involved. Thus, if a worker’s ability drops 1 level (eg, due to NP impairment), this presumably translates to thousands of jobs he/she can no longer perform.

The COMPASS batteries consist of multimodal, criterion-referenced instruments designed to establish participant skill levels in areas related to vocational functioning. They include computerized subtests and noncomputerized mechanical tasks that correspond to the previously mentioned DOT job levels. The computerized tests include vocabulary, reading, spelling, mathematics, language development (editing), problem solving, short-term visual memory, shape discrimination, size discrimination, and placing and tracking. The mechanical tasks involve alignment and driving, machine tending, and wiring. Raw scores from these tests were then converted into ability levels for each of the DOT classifications and summed in order to calculate a Work Assessment Total Score to provide a global index of vocational ability.

It is also possible to use only the computerized assessment components of the COMPASS. This abbreviated assessment, referred to as COMPASS-Lite, uses only a PC-compatible computer and collects no work samples requiring props. Thus, this shortened version can be completed in less than 30 minutes with hardware that is widely available in most research and clinical sites.

Previous research has found that even in people with subtle NP impairments, the COMPASS system is quite sensitive to deterioration in vocational skills. For instance, Heaton et al50 found that HIV+ patients with subtle NP deficits had considerable impairments in their performance on the vocational potential examinations. Further, within the NP-impaired group, there was a significant difference between their current level of vocational potential, measured with the performance-based tests, and their history of occupational functioning, with this discrepancy considerably larger than that seen in the unimpaired group. In another study using COMPASS,51 27 unemployed HIV+ patients were compared
on cognitive and illness-related factors to 27 unemployed schizophrenia patients and to 27 employed HIV patients. Employed HIV+ subjects were younger and better educated than were the unemployed schizophrenia and HIV+ subjects, who did not differ significantly on these factors. Both the HIV+ groups had a shorter duration of illness than did the schizophrenia group, and of the 2 unemployed groups, the schizophrenia subjects had been unemployed longer than the HIV+ subjects (8.7 vs 3.8 years). The 3 samples were compared on highest occupationally indexed vocational functioning, current vocational capacity (measured by COMPASS), and decline in their current functioning from their highest level of employment. The 2 unemployed groups did not differ in highest and current vocational abilities, and both were more cognitively impaired and performed more poorly on the COMPASS assessment than the employed HIV+ patients. When decline was examined, the unemployed schizophrenia patients had decline that was equivalent to that of the unemployed HIV+ patients. Most importantly, the schizophrenia patients showed evidence of greater impairment in occupational potential, in that one-third of the patients were unqualified for any of the 13 000 jobs in the index. Further, on average, the unemployed schizophrenia patients were rated as qualified for 479 jobs, while the unemployed HIV patients qualified for 872 and the employed HIV patients for 1506.

There are no treatment outcome data available yet for the COMPASS, but the test has many of the characteristics that make it quite a desirable instrument. It has high test-retest reliability and has excellent criterion-referenced validity.

**Summary and Conclusion**

Table 1 presents a summary of the characteristics of the different performance-based functional measures in terms of time required to complete the levels of training of the assessor and the number of props required for completion. This table is not intended for direct comparison of the different measures, which tap different domains of functional capacity in most cases, but for information's sake.

Performance-based tests of functional capacity offer advantages over other less direct methods of examining functional outcome. It will be important to continue to investigate the psychometric properties and sensitivity to change of performance-based tests. The inclusion of these assessments in clinical trials is an important step in linking treatment to functional outcome. It is often assumed that NP tests account for sufficient variance in functional outcomes that it is essential to include them in any comprehensive battery. However, as illustrated above, they do not always provide useful information about community behavior or change in community behavior, and functional assessments may be more useful in some circumstances. NP functioning is, at best, a mediator of community functioning, and NP tests are only indirect references of NP functioning. It can be assumed that some aspects of community functioning are less dependent on (mediated by) NP functioning than others. Hence, NP tests may not be useful for assessing outcomes in all clinical trials. For example, they are certainly more relevant for assessing the outcome of cognitive remediation interventions or medication trials that are expected to impact on neurocognition than they are for supported employment or social skills training. Research is needed to determine when NP tests account for sufficient variance to serve as useful outcome variables, when NP and functional outcome measures can be useful partners to give a broader perspective on outcome, and when functional measures can suffice. It is important that the field not be seduced into reifying NP measures based on their convenience and their current cachet and to consider that
more direct measures of functional disability are available for use in treatment studies.

References


35. Dickinson D, Bellack AS, Gold JM. Social/communication skills, cognition, and vocational functioning in schizophrenia. *Schizophr Bull.* In press.


37. Patterson TL, Goldman S, McKibbin CL, Hughes T, Jeste DV. UCSD performance-based skills assessment: development of


