A Video Ethnography Approach to Assessing The Ecological Validity of Neurocognitive and Functional Measures in Severe Mental Illness: Results From A Feasibility Study

Elizabeth Bromley*,1,2, Lisa Mikesell1,2, Andrea Mates3, Michael Smith3, and John S. Brekke4

1Department of Psychiatry and Biobehavioral Sciences, Semel Institute, Center for Health Services and Society, University of California, Los Angeles, CA; 2Desert Pacific Mental Illness Research, Education and Clinical Center, West Los Angeles VA Healthcare Center, Los Angeles, CA; 3Department of Applied Linguistics, Neurobiology of Language Research Group, University of California, Los Angeles, CA; 4School of Social Work, University of Southern California, Los Angeles, CA

*To whom correspondence should be addressed; tel: 310-794-3734, fax: 310-794-3724, e-mail: ebromley@ucla.edu

Background. The ecological validity of neurocognitive and functional measures in severe mental illness is poorly understood because of a lack of validated research methods to study community life-as-lived. We describe the development of a video ethnography method that measures naturalistic behaviors with codes called community performance indicators (CPIs). The method could provide a strategy to test the ecological validity of neurocognitive and functional assessments. Methods. We gathered up to 18.5 hours of video ethnography data on each of 9 subjects with schizophrenia selected for high or low composite scores on the MATRICS Consensus Cognitive Battery (MCCB). We used video ethnography to capture subjects’ everyday behaviors in their usual environments. We established 4 CPIs that showed excellent inter-rater and promising test-retest reliability: (1) behavioral activity level, (2) goal pursuit, (3) social interaction, and (4) problem solving. Results. (1) High and low MCCB subjects showed statistically significantly differences on all 4 CPIs. (2) MCCB composite scores were correlated with all 4 CPIs (r = .54 to −.77, P < .01 to .07). (3) The MCCB domain scores demonstrated some specificity in their correlations with the CPIs; eg, verbal learning, reasoning/problem solving, and social cognition were correlated with CPI domains of social interaction and problem solving. Conclusions. We present a method for reliably measuring everyday functional performance in schizophrenia. Results from a small select sample suggest that CPIs capture skills associated with neurocognition, supporting their use in a larger study of ecological validity.

| Key words: schizophrenia/neurocognition/functional performance |

Introduction

Deficits in independent living skills, occupational attainment, or social functioning account for a substantial portion of the disability associated with schizophrenia. Despite this, very few studies have directly observed and measured everyday functioning in naturalistic contexts.1 In addition, while numerous studies explore the role of cognitive deficits in schizophrenia, almost no studies have compared standardized assessments of cognition with assessments of what patients do in their daily lives in community settings.2 This means that we lack a fundamental understanding of whether and in what ways our best measures of neurocognition and functioning are associated with the daily living experiences of individuals with schizophrenia.

The need to establish the validity of these measures has become an area of considerable concern.3 While efforts are underway to compare the relative utility of different outcome measures,4 the ecological validity of these measures is poorly understood. Ecological validity concerns the predictive relationship between performance on a set of tests and behavior in real-world settings.5 Ecological validity is usually assessed by designing tests that mimic real-life scenarios, but a recent review concluded that “Observation in naturalistic settings is … the most robust approach because it allows the rater to evaluate whether the skills are actually implemented in the environment. Such data can also provide a measuring stick by which to evaluate the validity of other measures of everyday functioning.”6 Naturalistic observation “is technically the ideal”7 (p813) because it shows how behaviors unfold in the real world and in real time. Nevertheless, studies using direct observation in schizophrenia are rare.6

Video ethnography is a naturalistic nondirective method for recording daily activities as they occur in
real time and in an individual’s usual environments. In this feasibility study, we used video ethnography to provide a window on everyday functional performance and the manifestations of cognitive deficits in schizophrenia. The study aimed to (1) develop a video ethnography approach for capturing everyday community behavior, (2) assess its feasibility and tolerability for individuals with schizophrenia, (3) use video ethnography data to establish reliable measures of everyday behaviors, and (4) use the measures of observed behavior to obtain a preliminary signal on the ecological validity of neurocognitive tests in a small sample of individuals with schizophrenia.

**Functional Capacity, Performance, and Outcome**

Notable advances have been made in the conceptualization and measurement of neurocognition and functioning in schizophrenia. The MATRICS initiative established the validity and acceptability of a battery, the MATRICS Consensus Cognitive Battery (MCCB), to assess core neurocognitive domains relevant to the treatment of schizophrenia. In addition, a body of conceptual work on functioning has suggested that there are 3 critical functional domains: capacity, performance, and outcome. Functional “capacity” refers to an individual’s ability to perform tasks of daily living (eg, holding a conversation, preparing a meal). Capacity is most often assessed in controlled settings using role-play simulations. Functional “performance” refers to whether the individual engages in the above-mentioned behaviors in their natural living environments. Functional “outcomes” are the result of both capacity and performance and are typically measured as a level of achievement in work, independent living, and social domains. Research on functioning in schizophrenia provides scant insight about everyday functional performance. As a result, the pathway from functional capacity to performance and outcome is poorly understood. The notion of ecological validity provides a way to conceptualize and address these concerns.

**Ecological Validity**

Ecological validity concerns the relationship between an individual’s performance on a test and their real-world behaviors. Ecological validity is critical to ascertain because “…it is difficult to promote a putative indicator of functional status unless it predicts important aspects of patients’ actual behavior and adjustment in real-life contexts.” Because measures of functioning and neurocognition are increasingly used to assess the efficacy of treatments and to guide intervention design and testing, ecological validity has become an important area for research.

Establishing ecological validity is a process requiring a variety of methodologies and data collection techniques. There are 2 general approaches to ecological validity: “verisimilitude and veridicality.” Verisimilitude approaches are more common and aim to construct tests that mimic the tasks an individual might perform in life. Tests may assess individuals’ ability to perform skills in simulated settings (eg, a hospital kitchen), within role-plays, or with props that replicate everyday tools (eg, a bus map). Veridicality approaches, in contrast, assess the degree to which existing tests are empirically related to everyday behaviors and use the real world as the ecological standard. Both direct observation and experience sampling methods are veridicality approaches.

Veridicality and verisimilitude allow different types of inferences. Verisimilitude approaches yield data that can show strong correlations with other tests, such as cognitive assessments, but they do not allow inferences about everyday behaviors because assessment strategies have not been compared with performance in vivo. Limitations to the external validity of verisimilitude approaches may come from the tasks assessed, the testing setting, or the subject’s ability to engage in role-play. Veridicality approaches allow inferences about the relationship between test scores and naturalistic behaviors. The use of veridicality approaches to ecological validity has been extremely limited. Chaytor et al in their review did not find a single study in normal or clinical samples that compared neurocognitive measures with naturalistically observed behavior in community settings. We are aware of only one such study in schizophrenia. This study was limited in its use of general descriptors of behavioral styles, a single observer, and the use of field notes as the observational data.

**Video Ethnography**

Video ethnography is a nondirective method that allows the researcher to document subjects’ behavior while accompanying them in everyday settings. The ethnographer carries a camera and microphone, recording continuously while with the subject. The ethnographer interacts with subjects naturally and is careful neither to interfere with subjects’ behavior nor to comment on subjects’ plans. Clinical researchers have used video techniques for decades, yet video ethnography has not been used in schizophrenia. The current study aimed to develop the video method and to obtain a preliminary signal on the ecological validity of a neurocognitive battery in schizophrenia.

**Methods**

**Design**

We selected individuals diagnosed with schizophrenia living in the community. Subjects had completed a diagnostic interview and a neurocognitive test battery as part of an ongoing protocol. After selection and consenting, they were videotaped in their natural living contexts for up to 18.5 hours across multiple sessions. Taping was both done by PhD (E.B.) and graduate level
anthropologists (L.M., Gail Fox Adams, Brian Ellis, A.M.) who were blind to the neurocognitive scores of the subjects.

Sample Selection
Subjects with diagnoses of schizophrenia or schizoaffective disorder on the Structured Clinical Interview for DSM Disorders (SCID) were approached for enrollment if their MCCB composite score fell within the top or bottom one-third of all subjects (n = 155) in the ongoing study. Subjects hospitalized for psychiatric illness in the prior 6 weeks, with substance abuse within the past 6 months, on probation or parole, or with a comorbid cognitive disorder were excluded. Chart reviews and clinician interviews were used to verify diagnoses and the absence of active substance use prior to enrollment.

We enrolled 9 subjects (see table 1). Five scored in the bottom one-third and 4 scored in the top one-third of the cohort’s MCCB composite scores. There was a statistically significant difference between the 2 groups of subjects in the expected direction on the MCCB composite. Of the 9 subjects, 8 were African Americans, 1 was Asian, and 2 were female. Study samples from this clinic site are typically 55%–60% nonwhite, and because previous cross-sectional studies at this site showed few differences in neurocognitive or functional variables by race/ethnicity, we did not consider race in the recruitment protocol. Eight individuals had SCID-diagnosed schizophrenia, 1 had SCID-diagnosed schizoaffective disorder. Study procedures were approved by the institutional review board of the University of Southern California.

Neurocognitive Measures
We used 2 types of neurocognitive variables from the MCCB: the composite score and the 7 domain scores (Attention & Vigilance, Working Memory, Visual Learning, Verbal Learning, Reasoning & Problem Solving, Speed of Processing, and Social Cognition). The testing was done by a trained assistant supervised by a neuropsychologist involved in the development of the MCCB.

Camera Setup
We used a high-definition camera that weighs 8 pounds, measures 14” × 6” × 6”, and is carried with a shoulder strap while filming. A directional microphone was used along with a lavaliere omnidirectional microphone that was pinned to the ethnographer.

Filming Technique
We trained ethnographers to use a filming technique that is nondirective and naturalistic in order to maximize the ecological validity of the data. The technique was nondirective because subjects were told only that the ethnographers were interested in knowing more about their everyday routines. Ethnographers did not express preferences about plans or decisions. Ethnographers instructed subjects only to do “whatever you’d do if I were not here.” Naturalistic aspects of the technique included that ethnographers carried the camera discretely at their side, at the hip or under the arm, never looking through the eyepiece. They behaved in a relaxed friendly manner as a peer would do.

Duration of Videotaping
Ethnographers, not subjects, chose the time and date of observation sessions, though subjects’ preferences were accommodated. We taped individuals (n = 9) for from 1 to 6 sessions (average 4.4) during different times of week and weekend days. One subject was lost to both the clinic and the study after the first taping session but was located several months later to complete a tolerability survey. Sessions lasted 2–5 hours. Total taping time per subject (n = 9) ranged from 3 to 18.5 hours (average 10.3 h). A total of 93 hours of video were gathered.

Results
Tolerability, Reactivity, and Feasibility
Tolerability of the filming technique was measured using Likert scale items and interview questions administered after the second and final filming sessions. Asked if they felt uncomfortable being videotaped in public, 7 of 9 subjects disagreed (2) or strongly disagreed (1) after the second session; 7 of 7 did so after the last session (n = 16, average 1.6). Seven of 9 subjects agreed (4) or strongly agreed (5) that they forgot they were being videotaped after the second session and 5 of 6 did so after the last session (n = 15, average 3.7).

Reactivity was minimized because subjects were not informed of the data analytic strategy; they did not know how to appear skilled and could not disguise deficits, thus, video ethnography captured behavior as it naturally unfolded. In open-ended responses, 8 of 9 subjects described their behavior as “the same” or “no different.”
Table 2. Activities Observed in 1 Hour (Session 1, Hour 2), With ARS and GOL Scores, From 5 Subjects

<table>
<thead>
<tr>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
<th>Subject 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putting beans</td>
<td>Watching through</td>
<td>Sitting on bus to home</td>
<td>Watching TV</td>
<td>Walking to doctor’s appointment</td>
</tr>
<tr>
<td>in pot for soup</td>
<td>house to get coffee</td>
<td>walk home</td>
<td>goes to patio</td>
<td>Sits on bus bench</td>
</tr>
<tr>
<td>Returns to room</td>
<td>Greets peers</td>
<td>mail</td>
<td>to sit in sun</td>
<td>rests</td>
</tr>
<tr>
<td>Watches TV Goes</td>
<td>Checks</td>
<td>Places pot of</td>
<td>Attempts to</td>
<td>Resumes walking</td>
</tr>
<tr>
<td>downstairs to</td>
<td>whether house computer</td>
<td>water</td>
<td>make phone call</td>
<td></td>
</tr>
<tr>
<td>kitchen to check</td>
<td>in use</td>
<td>rice on stove</td>
<td></td>
<td></td>
</tr>
<tr>
<td>on beans</td>
<td>Drinks</td>
<td>Does dishes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>doorbell, directs</td>
<td>coffee &amp; smokes cigar</td>
<td>Empties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>visitor down hall</td>
<td>Converges with Approaches</td>
<td>Smokes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>to room, Watches TV</td>
<td>staff to take medication</td>
<td>cigarette</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Converges with staff about</td>
<td>check rice on stove</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ARS Highest

<table>
<thead>
<tr>
<th>GOL Domain</th>
<th>Daily life management, pursued (12 min); leisure, received (48 min)</th>
<th>Daily life management, pursued (46 min); leisure, received (14 min)</th>
<th>Leisure, received (60 min)</th>
<th>Daily life management, pursued (22 min); health, pursued (38 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>2, 1, 3, 3</td>
<td>0, 3, 3, 1</td>
<td>0, 0, 0, 2</td>
<td>2, 0, 2, 3</td>
</tr>
<tr>
<td>Subject 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: ARS, Activity Rating Scale; GOL, Goal Pursuit.

during filming (“I behaved the same as I always behave,” “basically we were just normal”); the ninth reported “being more polite.” Three subjects described participating as “exciting” and 6 described it as “easy” (“all easy,” “not really a big deal,” “it didn’t bother me at all”). Those subjects who perceived novelty in the taping situation reported this wore off after a short period, either in the first 30 minutes or after the first session. All subjects would recommend participation to others. These results mirror reactivity findings in other video-based research.

The discrete method made video recording feasible in public settings. No one with whom the subject interacted required consent because data analysis focused only on the subject. The public took little note of the camera. Peers at residences occasionally made comments but could be redirected (“they’re making a movie about me”). Subjects helped researchers obtain prior permission to tape in 2 locations (a classroom and a shelter). On 3 other occasions, a Study Information Letter was given to staff at residences in response to inquiries about the camera. The permission and information letters contained no reference to the subject having a mental illness. Before filming began, subjects were offered the opportunity to decline filming in any setting of their choice, but no subject did so.

Code Development

One purpose of this study was to develop a measurement strategy, or a set of codes, that would yield reliable descriptions of cognitively and functionally relevant everyday behaviors as seen on video. We used a team-based iterative process to develop quantitative codes for the filmed behavior. The team was blind to subjects’ neurocognitive scores. It included researchers with experience developing coding strategies for video (L.M., A.M., M.S., and Brian Ellis). The team began by reviewing paradigms used to assess cognition and functioning in schizophrenia (eg, testing batteries, measures of functional outcome, and capacity). An author (E.B.) interviewed 4 senior neuropsychologists who had differing opinions about the behaviors that correspond to cognitive domains, which highlighted the challenge of identifying relevant behavioral indicators of neurocognition. Then, the team studied strategies for decomposing and scoring complex behaviors, problemsolving therapy manuals, and diary methods for functional assessment. Each activity provided ideas about the categories of behaviors (eg, problem identification, goal pursuit, planning ahead) that might reflect everyday manifestations of cognitive and functional skill. Then, the team began a collaborative review of video data to consider which relevant behaviors could be seen and rated from video. Most domains of cognition (eg, visual learning and memory) could not be observed. Other cognitive tasks (eg, making change) were very rare or too varied to rate for their skill level. Table 2 provides sample activities from 1 hour of video.

The final codes share 4 features. The codes score skills rather than deficits because we could not reliably identify the absence of necessary behavior. The codes score only actions observed (not reported) on video because subjects’ reports (eg, “I’m going to class tomorrow with my brother.”) could not be validated. Unambiguous
measures (eg, duration) were used whenever possible to maximize reliability. Interactions with the ethnographer were excluded from codes.

**Community Performance Indicators**

The developed codes, called “community performance indicators” (CPIs), measure 4 domains of everyday behavior: behavioral activity level, problem solving, social interaction, and goal pursuit.

**Activity Rating Scale.** The Activity Rating Scale (ARS) measures behavioral activity level, which has been associated with clinical status in schizophrenia. The scale is based on time budget approaches and a diary rating scale that scores activity complexity and effort. The ARS scores 15-minute blocks of video. The 5-point (0–4) scale assigns higher scores for effortful behaviors, sequencing of behaviors, and multitasking. Active and motivated time blocks are scored higher than passive and solitary time blocks. Time in which a subject cooks a meal from scratch while fielding a phone call receives a “4”; time in which a subject watches TV receives a “0.” Each 15 minutes is scored for the highest ARS level achieved in that segment.

**PRO.** Problem-solving skill is measured with The Problem Solving (PRO) code. The PRO code is meant to reflect skills in executive functioning and reasoning and problem solving, which are core cognitive problems in schizophrenia. Discrete skills used repeatedly by subjects (eg, identifying opportunities) seemed to correspond to these domains. Inductive characterization of these behaviors as well as comparison with neurocognitive testing paradigms and principles of problem-solving therapy were used to refine the code. PRO behaviors are those in which subjects manage a novel circumstance, satisfy a need, plan ahead, or fix complications. PRO behaviors require set-shifting, problem identification, or positive problem orientation and include anticipating a peer’s needs, figuring out how to turn the oven on, and planning whether there is time to stop at the store. Each PRO behavior is coded on a 3-point scale assessing skill. PRO skill scores are higher for forward-looking and creative behaviors than for rote problem-solving approaches. Table 3 provides the rating scale for PRO skill.

**INT.** Social interaction is measured with The Social Interaction (INT) code, which rates the quality and quantity of interactions seen on video. Reported interactions and those with the ethnographer are excluded. Each social interaction is coded for duration and level of relatedness. “Duration” is measured in seconds. “Relatedness” is a 3-point scale based on the Maryland Assessment of Social Competence. Use of the interlocutor’s name, follow-up questions, eye contact, reactive affect, gestures, and intonation increase the relatedness score.

---

**Table 3. Sample Coding Protocol: PRO Skill Rating Scale**

| 1: Rote and rudimentary problem-solving skills | Subject forgets something or runs into an obstacle and fixes it (eg, “I forgot my keys; I’ll go get them.”). Subject runs into an obstacle and offers a rote solution without investigating if it is the only one (eg, “I think it’s okay for you to park there.”). Subject addresses an obvious need in a rote way, such as making simple offers about the present moment (“watch your step.” “we’ll cross the street now.” holding the door). The subject is providing information as requested (I think he’s upstairs), answering questions straightforwardly, or responding to others without elaboration. Subject follows another’s instructions to manage the problem (eg, ethnographer: “I think we can get the bus over there,” “I’ll tell the staff that your ears are ringing”), or depends on another to solve the problem. |
| 2: Anticipatory and additive problem-solving skills | Subject anticipates, looks ahead, and/or prepares (eg, “We’ll be getting off at the next stop.” “When will you be coming back?”). Subject may be responding to others (as in 1) but then also offers help that the subject did not have to offer or anticipates another’s need. That is, the subject may offer extra information, provide more than one solution (“We have splenda, and regular sugar, and stirrers over here”), or consider more than one solution. In this, the subject may seem to go above and beyond the minimum or show extra consideration for another. Overall, the subject is proactive or ahead of the situation at hand, rather than routinely responding to the moment. Subject turns a circumstance into an opportunity for improving his/her situation (eg, uses a bus stop to sit when subject needs to rest). |
| 3: Complex and creative problem-solving skills | The subject must show creativity and/or flexibility in offering, evaluating, and/or considering solutions. Subject weighs multiple strategies. Subject offers extra help (as in 2) but also offers help of more than one kind or persists in providing or reaching a solution. Subject must consider more than one kind of solution and/or multiple suggestions to manage a single problem. Subject demonstrates elaborate preparation, such as setting a specific and relevant action plan or organizing multiple phases of a project. |

---

**Note:** PRO, Problem Solving.

**GOL.** Goal pursuit is measured with The Goal Pursuit (GOL) code, which tracks time spent on discrete daily activities and behaviors. Based on the Activity Pattern Inventory and related research, one of 6 labels (socializing, leisure, spirituality, daily life management,
health, and vocation/education) are assigned to blocks of time. Each block is further rated as “pursued” or “received” to denote whether subjects passively take part or actively seek the activity. Time during which a subject talks with peers at her group home receives a GOL code of “received socializing.” Compared with the ARS, GOL codes longer blocks of time, captures the type of functional activity, and clarifies whether subjects actively pursue certain experiences.

We designed these CPIs to measure functionally and cognitively relevant behaviors. After data collection was complete, we asked 6 subjects to rate their attitudes toward the behaviors scored with the CPIs on a 5-point scale (strongly disagree to strongly agree). All 6 strongly agreed that it is important to them to set goals and pursue them, that it is important to them to make plans for how to spend the day, and that being able to solve problems is important to their daily lives. They also verified the importance of being active and interacting socially with others. This suggests that the CPI domains had face validity to our subjects.

**Coding Procedure**

All video ethnographers and coders were blind to subjects’ MCCB scores. Videos were digitized and viewed in InqScribe, a coding program that allows codes to be inserted (time-stamped) at exact locations in the video. Each CPI was carefully described in a codebook which was continually referenced during coding.

**CPI Variables**

Through the code development process described above, we defined the 4 CPI domains and the video data that would be coded with each. We focused on 6 CPI variables, at least 1 from each CPI domain. Each of these 6 variables was selected as a starting point because they appeared to represent core aspects of the data coded for each domain. In the PRO domain, the variable was the level of problem-solving skill averaged across all problems. In the GOL domain, the variable was the percent of total time spent on pursued goals. In the INT domain, there were 2 variables, 1 measuring the quantity of social interactions (the average duration of all social interactions) and the other measuring the quality of social interactions (the average relatedness of all interactions). For the ARS domain, we focused on the average of the highest ARS score across all time segments. We also calculated the SD of the ARS highest score across sessions, reasoning that this variable would reflect the range of behavioral activity across sessions, with greater range being an indicator of higher functioning.

**CPI Variables: Reliability**

Inter-rater reliability was established in 2 steps. First, a consensus rating among 3 code developers (PhD candidate or higher) was obtained for each CPI variable using several tape segments. Second, all coders (Master’s or BA level) were trained to a reliability criterion using the segments. After training, the raters reached high levels of inter-rater reliability on all variables (see table 4). Trained coders then coded all video.

Test-retest reliability was assessed to measure the stability of the CPI variables from one observation session to the next. We saw test-retest reliability as a reasonable, but not essential, criterion for each variable as we expected to need more than a single observation session to comprehensively document functional performance. We tested internal consistency over 4 observation sessions using Cronbach’s alpha. Internal consistency was above .59 on the CPIs except for the social interaction duration variable (see table 4). Collapsing longer time periods (sessions 1 + 2 vs 3 + 4) improved test-retest reliability of the social interaction duration variable. This suggests that more than a single observation session is needed to assess aspects of social interaction in naturalistic settings.

**Relationship Between CPI Variables and Neurocognitive Variables: Ecological Validity**

We examined the relationship between our 6 CPI variables and the MCCB variables. Two of the 6 variables showed no statistically significant relationships with MCCB scores: the average ARS highest score and average relatedness in social interactions; these variables are not reported on further. The analyses reported below include 1 variable from each CPI domain: average problem-solving skill, average duration of social interactions, percent time spent in goal pursuit, and the SD of the average ARS highest score.

Based on the assumption that the CPIs reflect natural behavior and performance, several sets of findings provide a preliminary signal on the ecological validity of the MCCB. First, the high and low neurocognition subjects showed statistically significantly differences on all 4 CPI domains in the predicted direction (better neurocognition corresponding to better functioning) (chi-square, \(df = 1\) 3.0–6.0, \(P < .05\), one-tailed, in all cases using
Table 5. Community Performance Indicators: Pearson Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>ARS Highest SD</th>
<th>GOL % Pursued</th>
<th>INT Duration</th>
<th>PRO Average Skill</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of processing T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>-.082</td>
<td>.533</td>
<td>.004</td>
<td>-.039</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.423</td>
<td>.070</td>
<td>.496</td>
<td>.464</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Attention/vigilance T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.539</td>
<td>.674*</td>
<td>.172</td>
<td>.327</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.084</td>
<td>.023</td>
<td>.329</td>
<td>.214</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Working memory T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.857**</td>
<td>.578</td>
<td>.110</td>
<td>.730*</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.003</td>
<td>.051</td>
<td>.389</td>
<td>.020</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Verbal learning T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.280</td>
<td>-.146</td>
<td>.907**</td>
<td>.465</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.251</td>
<td>.354</td>
<td>.000</td>
<td>.123</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Visual learning T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.497</td>
<td>.361</td>
<td>.389</td>
<td>.741*</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.105</td>
<td>.170</td>
<td>.150</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Reasoning and problem solving T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.408</td>
<td>.294</td>
<td>.797**</td>
<td>.692*</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.158</td>
<td>.221</td>
<td>.005</td>
<td>.029</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Social cognition T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.266</td>
<td>-.191</td>
<td>.419</td>
<td>.641*</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.262</td>
<td>.311</td>
<td>.131</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Composite T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.640*</td>
<td>.537</td>
<td>.557</td>
<td>.772*</td>
<td></td>
</tr>
<tr>
<td>Significant (one-tailed)</td>
<td>.044</td>
<td>.068</td>
<td>.060</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Note: Abbreviations as described in notes in Tables 2, 3, and 4. *Correlation is significant at the .05 level (one-tailed). **Correlation is significant at the .01 level (one-tailed).

Kruskal-Wallis test statistic). Second, the Pearson correlations between MCCB composite scores and the 4 CPI variables ranged from .54 to .77 (P < .01 to .07) (see table 5). Third, the Pearson correlation between the MCCB composite and a z-score CPI composite was .82 (P < .01). Fourth, the CPI domains showed a pattern of specific correlations with neurocognitive domains. Working Memory and Visual Learning had strong associations across the CPI domains. Verbal Learning, Reasoning and Problem Solving, and Social Cognition had strong associations with CPI domains of social interaction and problem solving, while Attention-Vigilance and Speed of Processing were associated with the CPI domain of goal pursuit (r > .42 or P < .05 in all cases). Two scatterplots of CPI variables and the composite MCCB score are provided for visual inspection (see figure 1). These findings are very similar when using a nonparametric correlation coefficient. Neither BPRS scores nor BASIS-32 scores showed significant correlations with the CPIs.

Relationships Among the 4 CPI Variables

The Pearson correlations among the 4 CPI variables ranged from .07 to .72 (average r = .45). The Cronbach’s alpha among the 4 variables is .79. These findings suggest that the CPI variables can be used separately, and that a composite score is also viable.

Discussion

This article describes the development and pilot testing of a video ethnography method to measure functional performance in schizophrenia. Four domains of everyday behaviors (behavioral activity level, goal pursuit, social interactions, and problem solving) were reliably coded with CPIs. CPIs demonstrated adequate inter-rater and promising test-retest reliability. Groups with high and low MCCB scores differed significantly on all 4 CPIs. Variables from each CPI domain as well as a z-score CPI composite score showed correlations with composite MCCB scores. In addition, some MCCB domain scores showed specificity effects with regard to everyday behaviors (e.g., Verbal Learning, Reasoning and Problem Solving, and Social Cognition were most strongly correlated with CPI domains of social interaction and problem solving). If these preliminary results were to remain robust in a larger sample, it would suggest that the MCCB alone might provide reliable inferences about individuals’ everyday behaviors in usual environments.

This video ethnography method provides the most detailed and ecologically valid assessment of functional performance yet available. The ability to measure functional performance is significant for several reasons. First, functional performance is equivalent to the behaviors that constitute day-to-day functioning; as such, it may link functional capacity to functional outcome. A better understanding of the behaviors that comprise day-to-day functioning could clarify the mechanisms through which individuals achieve functional improvement. Second, an understanding of functional performance could identify environmental supports that assist individuals with functional achievement. These factors may moderate the impact of neurocognitive deficits, and they may be modifiable. Third, functional performance can provide a gold-standard measure against which structured or laboratory-based assessments of cognition and functioning can be validated. Finally, data on functional performance decompose complex behaviors like neurocognition and functioning into their component parts. This may provide useful information about the everyday manifestations of cognitive deficits, which are poorly understood. Such findings may also facilitate the translation of data on neurocognition to clinicians, patients, and families.
It may also aid studies on the biological and neurobiological underpinnings of schizophrenia.

This naturalistic observation method proved feasible, and subjects rated its tolerability high and its reactivity low. Safeguards used to maximize discretion and autonomy for subjects were effective. We experienced some difficulties with recruitment and retention, suggesting that these require careful attention in a follow-up study, as has been suggested by other video researchers. Reactivity to videotaping is understudied but appeared to be
minimal, as has been found in other studies.\textsuperscript{30,31,48} It is possible that subjects increased the amount or quality of their behavior during filming, but they could not do so in a manner that exceeded their usual capacity. With regard to the 4 CPIs, we believe they reflect significant aspects of functional performance for individuals with schizophrenia, but they are by no means exhaustive, and future research may identify additional domains for measurement. Moreover, our pilot work defined additional CPI variables with good inter-rater and test-retest reliability but no statistically significant relationship with neurocognition (eg, relatedness in social interactions, average highest ARS score). These variables should be revisited in a larger study.

In terms of limitations, this study included a very small sample that was selected purposively to maximize neurocognitive differences between groups of subjects. The sample was all ethnic minority and predominantly African American. While it is more common that studies lack adequate ethnic minority representation,\textsuperscript{49} the absence of nonminorities is an equal limitation. Sampling was blind to race and ethnicity because we believed that any differences on neurocognition\textsuperscript{27,50,51} or social functioning\textsuperscript{28} based on ethnic minority status would be minor and that any mean differences would not likely effect the relationship between neurocognition and functional performance. The results require replication in a more ethnically heterogeneous sample in order to understand any impact that race or ethnicity might have on the performance of the CPI variables. The impact of other variables, including age, gender, premorbid status, and treatment history also need to be examined in a representative sample. Given how the sample was selected, it is also unlikely that the magnitude of the correlations will be replicated in a larger more representative sample. Overall, the generalizability of the results requires further investigation. Finally, our preliminary signal on the ecological validity of the MCCB was based on an exploratory empirical approach that eliminated 2 CPI variables because they were not good predictors. We did find 4 strong CPI variables that reflected each CPI domain. These results are notable given that the everyday behaviors we were able to code were quite distinct from the skills directly assessed on a cognitive assessment battery. Overall, these variables form a solid basis for future work.

As developed here, video ethnography is a labor-intensive approach that is not yet feasible for use as a routine assessment of functioning. Instead, we see it as a promising strategy with which to assess the ecological validity of widely used neuropsychological and functional measures, an area of substantial importance. The test-retest reliability of 3 of the 4 CPI variables across sessions was adequate, and the composite CPI measure was correlated with MCCB composite scores, suggesting measures from brief observation sessions may be able to be pooled. These findings indicate possibilities for adaptations that use brief episodes of observation. However, exploration of brief adaptations requires that the psychometric characteristics of the CPIs be explored in a larger more representative sample.

**Funding**

VA Desert Pacific Mental Illness Research, Education and Clinical Center to E.B; University of California, Los Angeles Faculty Senate Council on Research to E.B.; University of Southern California Clinical and Translational Science Institute to J.S.B.

**Acknowledgments**

The authors wish to thank Brian Ellis, Gail Fox Adams, Lisa Davis, Victoria Valdes, Mark Morales, Jenebah Lewis, Betsy Phillips, Lily Zhang, and Patricia Lewis for invaluable contributions to this research. The authors have no financial conflicts of interest to report.

**References**

E. Bromley et al.


47. Bromley E. Barriers to the appropriate clinical use of medications that improve the cognitive deficits of schizophrenia. *Psychiatr Serv.* 2007;58:475–481.

