

# UC San Diego

## UC San Diego Previously Published Works

### Title

Utilizing Technology to Enhance the Ecological Validity of Cognitive and Functional Assessments in Schizophrenia: An Overview of the State-of-the-Art.

### Permalink

<https://escholarship.org/uc/item/36n961w0>

### Journal

Schizophrenia Bulletin Open, 5(1)

### Authors

Horan, William

Moore, Raeanne

Belanger, Heather

et al.

### Publication Date

2024

### DOI

10.1093/schizbullopen/sgae025

Peer reviewed

# Utilizing Technology to Enhance the Ecological Validity of Cognitive and Functional Assessments in Schizophrenia: An Overview of the State-of-the-Art

William P. Horan<sup>\*,1,2</sup>, Raeanne C. Moore<sup>3</sup>, Heather G. Belanger<sup>4,5</sup>, Philip D. Harvey<sup>6</sup>

<sup>1</sup>Karuna Therapeutics, A Bristol Myers Squibb Company, Boston, MA, USA; <sup>2</sup>University of California, Los Angeles, CA, USA; <sup>3</sup>University of California, San Diego, CA, USA; <sup>4</sup>Cognitive Research Corporation, St Petersburg, FL, USA; <sup>5</sup>Departments of Psychiatry and Behavioral Neurosciences, and Psychology, University of South Florida, Tampa, FL, USA; <sup>6</sup>University of Miami Miller School of Medicine, Miami, FL, USA

\*To whom correspondence should be addressed: Karuna Therapeutics, 99 High Street, Floor 26, Boston, MA 02110, USA; e-mail: [whoran@karunatx.com](mailto:whoran@karunatx.com)

**Cognitive impairment is a core feature of schizophrenia and a key determinant of functional outcome. Although conventional paper-and-pencil based cognitive assessments used in schizophrenia remained relatively static during most of the 20th century, this century has witnessed the emergence of innovative digital technologies that aim to enhance the ecological validity of performance-based assessments. This narrative review provides an overview of new technologies that show promise for enhancing the ecological validity of cognitive and functional assessments. We focus on 2 approaches that are particularly relevant for schizophrenia research: (1) digital functional capacity tasks, which use simulations to measure performance of important daily life activities (e.g., virtual shopping tasks), delivered both in-person and remotely, and (2) remote device-based assessments, which include self-administered cognitive tasks (e.g., processing speed test) or functionally-focused surveys regarding momentary activities and experiences (e.g., location, social context), as well as passive sensor-based metrics (e.g., actigraphy measures of activity), during daily life. For each approach, we describe the potential for enhancing ecological validity, provide examples of select measures that have been used in schizophrenia research, summarize available data on their feasibility and validity, and consider remaining challenges. Rapidly growing evidence indicates that digital technologies have the potential to enhance the ecological validity of cognitive and functional outcome assessments, and thereby advance research into the causes of, and treatments for, functional disability in schizophrenia.**

*Key words:* technology/enhance/ecological/validity/cognitive/functional

## Introduction

Since the mid-20th century, the predominant approach to cognitive assessment in schizophrenia has involved the administration of paper-and-pencil tasks by highly trained research personnel during in-person meetings conducted in a well-controlled clinical setting. This approach originated from early efforts to apply clinical neuropsychology tests, initially developed for diagnosing so-called “organic” cognitive impairment due to focal brain lesions,<sup>1</sup> to people with schizophrenia. Over the past 50 years, hundreds of schizophrenia studies have utilized these types of traditional tasks or their modifications.

Chief learnings from research using traditional tasks are that cognitive impairment associated with schizophrenia (CIAS) is substantial (typically 1–1.5 SD below healthy comparison norms), spans multiple cognitive subdomains, present from before the time of psychosis onset, relatively stable across the longitudinal course of illness, detectable at attenuated levels in unaffected biological relatives, and consistently correlated with and predictive of level of functional disability.<sup>2,3</sup> This corpus of research has been central to reconceptualizing of schizophrenia as a neurocognitive disorder.<sup>4,5</sup> The MATRICS Consensus Cognitive Battery (MCCB),<sup>6,7</sup> which is administered in-persons and is comprised of 9 paper-and-pencil tasks plus 1 computerized test that assess 7 cognitive subdomains, remains the gold standard outcome measure for clinical trials aimed

at improving CIAS. Unfortunately, despite major efforts, the goal of developing efficacious treatments for CIAS remains unfulfilled.<sup>8</sup>

While traditional neuropsychological tasks have provided significant insights in schizophrenia and beyond, advances in neuroimaging technology and our understanding of brain-behavior relationships have challenged some foundational aspects of early neuropsychology.<sup>9</sup> For example, theories of brain function localization evolved, revealing that performance on many traditional tasks often reflects multiple interacting neurocognitive circuits, with dysfunction at multiple possible locations resulting in impairments on the end-state “localized” task.<sup>10</sup> This historical shift in perspective has steered clinical neuropsychological away from diagnosing focal brain pathology toward developing tools with ecologically validity-tools that measure the real-world impact of cognitive challenges on patients’ ability to function in the real-world, such as being able to work, handle finances, drive, or live independently.<sup>11</sup>

From a theoretical standpoint, these approaches are consistent with ecological models of cognition, which emphasize understanding cognitive processes within the context of individuals’ everyday environments.<sup>12</sup> By utilizing these models, digital tools are able to capture the complexity and variability of real-world cognitive abilities, thus providing more comprehensive assessments reflective of real-life functioning.

### *Goals and Structure of This Review*

The goal of this narrative review is to provide an overview of digital technologies designed to improve the ecological validity of cognitive and functional assessments and their application to schizophrenia research. The review consists of 4 main sections. First, “Ecological validity: definition, methodological approaches, and relevance to schizophrenia” section provides background on ecological validity, the evolution of neuropsychological tests, and contemporary technology-enhanced assessments within this definitional framework. Next, “Digital functional capacity measures in schizophrenia” and “Portable device-based assessments in schizophrenia” sections take a deeper dive into examples of technology-enhanced ecologically focused approaches that are particularly relevant for schizophrenia research. “Digital functional capacity measures in schizophrenia” section examines digital functional capacity tasks, which use simulations to measure how well participants can perform important daily life activities, while “Portable device-based assessments in schizophrenia” section describes portable device-based assessments, including remotely administered mobile cognitive tasks, brief surveys about current activities and experiences, and passive data collection. Since the approaches covered in the second and third sections are relatively new, we aim to illustrate the range of emerging

new tools in these areas and describe the current evidence supporting their utility in schizophrenia rather than providing an exhaustive review. Finally, in “Conclusions” section, we take stock of the current state of these approaches, consider ongoing challenges, and describe key future directions.

### **Ecological Validity: Definition, Methodological Approaches, and Relevance to Schizophrenia**

Ecological Validity is commonly defined as “the functional and predictive relationship between a patient’s performance on a set of neuropsychological tests and their behavior in a variety of real-world settings.”<sup>13</sup> In this definition, the goal is indexing how testing results obtained in a separated, highly controlled artificial setting correspond to how people behave during the naturalistic course of their everyday lives in the community. For many years, critiques and debates surrounding traditional neuropsychological measures have focused on their limited ecological validity.<sup>14-17</sup> Proponents of using tasks with higher ecological validity argue that there is a need to better understand how patients behave in the real-world, which is essential for creating interventions that maximize functioning more effectively in everyday life.

As shown in [figure 1](#), the ecological validity of performance-based assessments is frequently evaluated in terms of 2 dimensions<sup>13,18</sup>:

1. Verisimilitude: “the degree to which the cognitive demands of a test theoretically resemble the cognitive demands in the everyday environment.”<sup>13</sup> This overlaps with the concept of “representativeness” and is based on judgement rather than any empirical indicators. Task verisimilitude ranges from low to high levels of stimuli /activities with features of real-life.
2. Veridicality: “the degree to which existing tests are empirically related to measures of everyday functioning.”<sup>13</sup> This overlaps with the concept of “generalizability” and is measured by correlation with other measures of functional outcome, which can be assessed by achievement of functional milestones (e.g., living independently, obtaining competitive employment), questionnaires, or clinical/informant ratings of real-world functioning. Task veridicality ranges from low to high levels of association with functional outcome or rehabilitation success.

The evolution of cognitive tasks since the mid-20th century can be depicted in terms of where they fall on these 2 dimensions.

### *Traditional Measures*

Traditional tasks from neuropsychology have relatively low verisimilitude. These tasks typically use stimuli that are simplistic, static, and neutral in emotional valence,

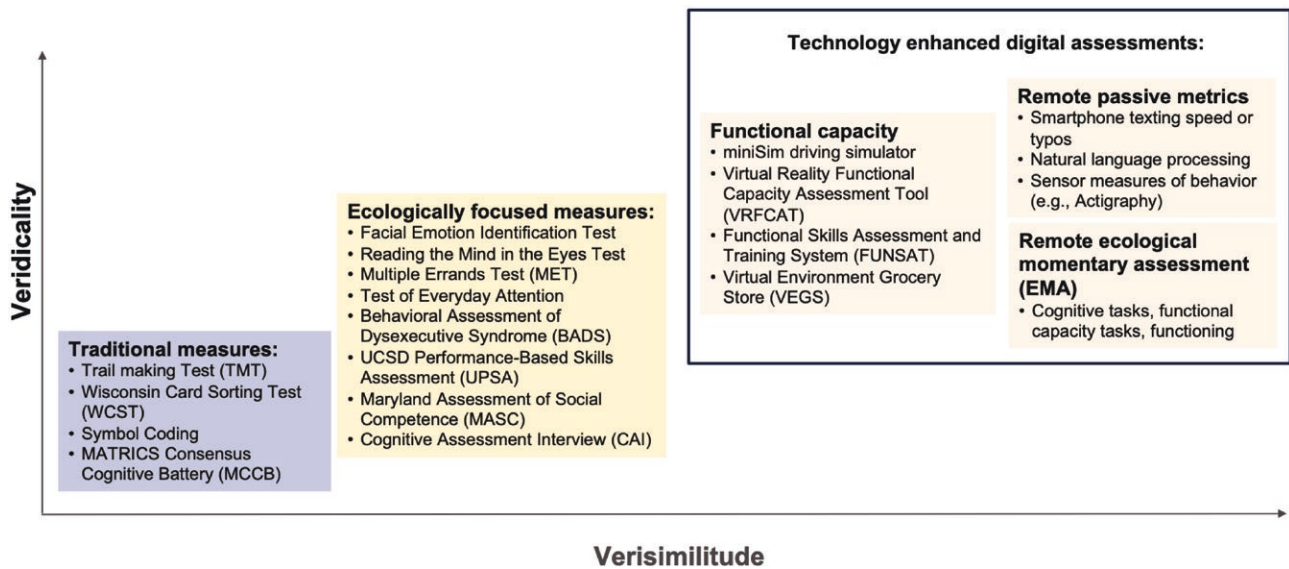


Fig. 1. Ecological validity of cognitive assessments.

along with well-defined instructions, to isolate cognitive constructs (e.g., executive functions, processing speed) under conditions that are as similar as possible across tests and individuals. Commonly used examples in schizophrenia include the Wisconsin Card Sorting Test,<sup>19</sup> which uses colored geometric shapes to assess conceptual set-shifting, and the Digit Symbol Coding Task,<sup>20</sup> which involves matching abstract symbols to numbers within a specified time period to assess psychomotor processing speed and working memory. The nature of these tasks differs from many of the complex cognitive challenges that we face in daily life, which are often poorly defined, open ended, require multi-tasking, and are imbued with social and emotional meaning as well as multiple potential distractions. Traditional tasks are amenable to standardized administration, scoring, and interpretation (e.g., with norms). Notably, many traditional paper-and-pencil administered cognitive tasks or test batteries have been adapted into computerized versions, which have several advantages in terms of computer-assisted administration to enhance standardization, automated scoring, and data management efficiencies, particularly for use in larger scale research and clinical trials.<sup>21,22</sup> However, because these computerized tasks are direct digital migrations of paper and pencil tasks and are designed to be administered in a controlled (aka “ideal”) environment, their stimuli and procedures, regardless of administration format, show relatively little direct relation to many of the important cognitive demands of daily life.

Although traditional neuropsychological tests were born out of efforts to diagnose brain injuries, and ecological validity was never a consideration in their development, these tests do show some degree of veridicality. However, reviews in the general neuropsychology literature suggest a varying degree of association with

everyday functioning across tasks, with many showing only a modest association to functioning.<sup>11,17</sup> Their ability to predict rehabilitation success from baseline scores is generally modest, with improvements on these tasks with training/treatment typically showing limited generalizability to real-world functional gains.

These findings mirror the results in schizophrenia. In general, performance on these measures generally show modest associations ( $r_s = 0.20-0.30$ ) with real-world functioning, accounting for small to moderate variance in outcome<sup>23</sup>; findings are stronger for composite scores that integrate multiple subtests. Further, low generalizability of improvements in standard cognitive remediation (without concomitant psychiatric rehabilitation) to real-world functional benefits has long been an issue in schizophrenia.<sup>24</sup> While functional outcome in schizophrenia is, of course, determined by multiple factors aside from cognition (e.g., motivation, psychosocial support, socioeconomic conditions, adjunctive psychiatric rehabilitation, and cultural factors to name just a few), the low verisimilitude and modest veridicality of the traditional neuropsychological tests suggest that we are not measuring all the “right stuff”<sup>23,25,26</sup> to optimally understand and improve functional outcomes in schizophrenia.

#### *Early Ecologically Focused Measures of Cognition and Functional Capacity*

Ecological tasks encompass a broad class of measures designed with the intention of more closely capturing the essence of everyday cognitive skills as they are used in real-life situations.<sup>11,18</sup> These tasks have higher verisimilitude in terms of their more obviously function-focused stimuli and direct relevance to real-world activities. For

example, early pencil-and-paper social cognitive measures, such as the Face Emotion Identification Test<sup>27</sup> or the Reading the Mind in the Eyes test,<sup>28</sup> incorporate realistic photos of social stimuli. Additional examples include the Test of Everyday Attention,<sup>29</sup> which assess attentional processes using more real-world stimuli and activities (e.g., searching a map, counting elevators), and the Behavioral Assessment of the Dysexecutive Syndrome,<sup>30</sup> which assesses planning and strategic thinking in relatively unstructured contexts (e.g., searching for keys, problem solving how to remove a cork from a tube while following certain rules).

In schizophrenia research, early ecological tasks have had a large impact on the field. This is particularly true for social cognition, which has become a major research topic due largely to the evidence that participants with schizophrenia show substantial impairment on pencil-and-paper or computerized social cognitive tasks (e.g., emotion perception, mentalizing) and that these tasks are more robust predictors of poor social functioning in the community than traditional nonsocial cognitive tasks.<sup>2,31–34</sup> These findings have inspired novel social cognitive intervention approaches aimed at enhancing functional outcomes.<sup>35–37</sup>

Other early paradigms assessed performance of real-world tasks in naturalistic settings or with props or role-play scenarios to simulate real-world activities. For example, the Multiple Errands Test<sup>38</sup> involves purchasing specific items, collecting and writing down specific information, and navigating to particular locations while following a set of rules in an actual shopping center. To address practical and scientific challenges (e.g., with standardized administration and scoring) of implementing these naturalistic paradigms in real-world settings, other tasks have used props or role-play tasks to assess ability to perform functional or social skills in a controlled experimental setting.<sup>39</sup> An example is the Executive Function Performance Test,<sup>40</sup> which evaluates performance in the execution of 4 tasks: simple cooking, telephone use, medication management, and bill payment. Despite considerable variability across ecological measures in the extent to which their relation to real-world outcomes has been tested, reviews in the general neuropsychology literature indicate that these tests often contribute additional variance to the prediction of real-world functioning beyond traditional tests.<sup>11</sup> It should be emphasized, however, that using ecological tools with higher verisimilitude does not guarantee stronger veridicality.

In schizophrenia research, an important class of early performance-based ecological tasks focuses on measuring “functional capacity.” This term refers to the ability to perform essential everyday living skills within controlled situations,<sup>41</sup> including everyday living activities, work skills, and social skills. Given the close linkage of functional capacity with cognitive abilities as measured by

traditional neuropsychological tasks,<sup>41,42</sup> some assessment of these domains is presently required (i.e., a co-primary requirement) by regulatory authorities to demonstrate the functional relevance of interventions targeting cognition with either pharmacological or device-based computerized cognitive training (CCT) strategies.<sup>43</sup> Two global assessment strategies have been proposed—structured interviews about real-world performance of cognitively demanding functional skills with patients and informants and performance on simulated assessments of functional skills.

Several interview-based measures have been developed, such as the Schizophrenia Cognition Rating Scale (SCoRS)<sup>44</sup> and the Cognitive Assessment Interview (CAI),<sup>45</sup> in which patients and informants are asked questions about the patient’s cognitive deficits and the degree to which these deficits impair their day-to-day functioning (e.g., cognitive challenges during everyday tasks such as conversations or using electronic devices). A major challenge with this approach is that retrospective self-reports of level of cognitive ability and functioning are notoriously inaccurate for most people with schizophrenia.<sup>46,47</sup> Although high quality informant ratings can be highly accurate, many patients do not have a willing and available informant who sees them with sufficient frequency to provide meaningful ratings.<sup>41,48</sup> Further, these interviews blend recollections about past performance with prospective judgments regarding abilities. As both healthy people and those with schizophrenia commonly misjudge their functional abilities in predictable (i.e., commonly overestimated) ways, this element of the assessment includes another challenge to validity.

Widely used simulation tasks in schizophrenia include the UCSD Performance-Based Skills Assessment (UPSA),<sup>49</sup> which was developed in the 1990s and involves using a series of props (e.g., grocery items, a telephone, bills) to perform daily tasks such as shopping and paying bills, or the Maryland Assessment of Social Competence<sup>50</sup> which involves role playing various type of interactions with a highly trained confederate (e.g., meeting a new colleague at work). While such measures avoid reliance on patient self-reports and demonstrate a closer association with cognitive test data than the interview-based measures,<sup>48,51,52</sup> they have several limitations, including substantial equipment requirements, large training demands for standardized administration and scoring, a lack of alternate forms and norms, the still artificial experimental context in which they are conducted (i.e., pretend that there is an emergency and call the correct number), and content that can be outdated or even obsolete. For example, the UPSA includes tasks such as using a landline phone to dial directory assistance or paying bills with a paper check, which are no longer used in many cultures (particularly by younger people). However, a new generation of functional capacity tasks is capitalizing on technology

to address the practical and scientific challenges of early tools and, hopefully, to maximize both verisimilitude and veridicality.

Advances in digital technology have had a massive impact on our lives in the 21st century. Test developers have embraced these advances and tapped into their potential to improve the ecological validity of cognitive and functional assessments, whether they are conducted within or outside of a clinical setting. Two technology-enhanced approaches with particular relevance to schizophrenia research are performance-based digital functional capacity measures, which provide increasingly realistic simulations of functional tasks for in-person assessments, and active and passive remote digital assessments, which enable cognitive and functional assessments to be conducted as people perform their regular activities in the community. The following 2 sections aim to illustrate the range of emerging new tools in these areas and describe the current evidence supporting their utility in schizophrenia.

### Digital Functional Capacity Measures in Schizophrenia

Regarding functional capacity measures, technology offers the possibility of creating ecologically valid simulation tasks with unprecedented levels of verisimilitude. These tools range from 2-dimensional, nonimmersive presentations on a computer or tablet screen, using graphics that have become increasingly sophisticated and realistic, to immersive 3-D and multisensory Virtual Reality (VR)-based tools.<sup>53</sup> These computer interfaces are uniquely capable of simulating a wide range of realistic virtual environments in a safe laboratory setting, with a richness that was not previously possible.

Applications of these simulations include both assessment and intervention/rehabilitation. Additionally, these software-based programs provide a high level of experimental control, including excellent measurement precision and standardization, the ability to easily update task requirements with changes in technology, and to create alternate forms for repeated testing, and error-free scoring and data management that can meet regulatory standards. Further, VR environments have the potential to enhance task engagement through gamification, hence increasing participant motivation, thereby increasing validity and compliance. As detailed below, a range of emerging technology-enhanced simulation tasks are relevant to schizophrenia, including functional assessments such as driving, banking, using the internet, meal planning, grocery shopping. Some of these tasks are now cloud-based and fully remotely deliverable, with systematic research on the quality of their migration from in-person to remote delivery.<sup>54</sup>

A substantial range of measures relevant to schizophrenia are currently available, though only a subset of them have been used in this disorder. This section describes 4 measures used to varying extents in schizophrenia. The

measures differ in several ways, including the type(s) of technology involved, whether they address 1 or multiple functional areas, and whether they simulate sequential daily activities (e.g., shopping) or performance of specific tasks on a more modular basis (e.g., using an automated teller machine touch screen).

### *MiniSim Driving Simulator*

Driving is an important functional task for people with schizophrenia given its association with autonomy and community inclusion and many people with schizophrenia do not drive.<sup>55</sup> The United States Food and Drug Administration (FDA) issued guidelines regarding the use of driving simulation in clinical trials,<sup>56</sup> which state that drugs with certain characteristics should be evaluated in terms of their potential effects on driving ability. The recommended means for testing driving ability are over-the-road (OTR) testing or driving simulation. Given the cost, safety issues, and other impracticalities of conducting OTR assessment, driving simulation represents a VR approach with both verisimilitude, and veridicality. The most recent reviews of the driving simulation literature<sup>57,58</sup> provide evidence supporting the validity of driving simulation vis-à-vis on-road driving.

Modern driving simulators are PC-based simulators that provide a realistic automotive experience with some type of visual display of the driving environment on a screen or several screens, as well as typically a driving cockpit with some kind with a steering wheel and pedals. The “gold standard” driving simulator is a simulator that is actually an instrumented car housed within a laboratory at the National Advanced Driving Simulator (NADS) at the University of Iowa, which has established validity.<sup>59,60</sup> The NADS provides a high degree of fidelity via features such as 360-degree viewing, vehicle-specific components with a 13-degree-of-freedom motion system, interactive traffic, and an actual vehicle cab. The NADS offers the highest degree of fidelity possible with a simulator, though the disadvantage is the cost and lowered efficiency associated with having an actual car cab enclosed inside a facility, as well as the engineers and other personnel that are needed to run it.

A more practical driving simulator for use in clinical trials would be the MiniSim,<sup>61</sup> consisting of a three-screen-wide display of the driving environment with 108-degree view of the roadway, a driving cockpit with a full-size regulation steering wheel and realistic pedals, and an actual automotive seat that meets current US National Highway Traffic Safety Administration standards (see [figure 2](#)). It has been validated against OTR driving tests<sup>59,62</sup> and has been used in various clinical trials,<sup>63-68</sup> though not in any published schizophrenia trials.

Only a handful of studies have used driving simulation with individuals with schizophrenia, largely with more rudimentary driving simulators than are used today.<sup>69-71</sup>



Fig. 2. miniSim driving simulator.

In these studies, individuals with schizophrenia tended to underperform relative to matched, healthy controls on most parameters, including standard deviation of lateral position (SDLP), which is the parameter frequently used in clinical trials with the most established validity.<sup>72</sup> However, more modern driving simulators have revealed no<sup>73</sup> or relatively small differences in those with schizophrenia, including slower speed and hinderance of surrounding drivers.<sup>74</sup>

Significant associations have been demonstrated between performance on cognitive tasks of concentration and vigilance and performance in a driving simulator<sup>75</sup> in those with schizophrenia. Additionally, some antipsychotic (or other commonly used) medications may impair reaction time, vigilance, and driving performance in individuals with schizophrenia.<sup>76</sup> While second-generation antipsychotics (SGA) typically have less adverse impact on these abilities than conventional antipsychotics,<sup>75-77</sup> severity of extrapyramidal symptoms among those taking SGAs correlates negatively with driving fitness.<sup>78</sup> Overall, like some of the other uses of driving simulation recommended by the FDA, this type of assessment may be best suited toward detection of worsening in performance as an adverse event of treatment in schizophrenia.

#### *Virtual Reality Functional Capacity Assessment Test (VRFCAT)*

The VRFCAT<sup>79</sup> is a semi-immersive simulation of real-world situations that is administered via a tablet interface. The VRFCAT was developed to measure an individual's capacity to perform independent activities of daily living (IADLs) in 4 different functional abilities: food preparation, using transportation, shopping, and managing currency. The VRFCAT's realistic, interactive and immersive environment consists of 4 mini scenarios that include navigating a kitchen, catching a bus to a grocery store,

finding/purchasing food in a grocery store, and returning home on a bus. As displayed in figure 3, these scenarios were developed using immersive "first-person" gaming technology. Patients complete the scenarios through a progressive storyboard design. There are 12 different tasks or "objectives," described in the Figure, and for each objective, the dependent variables are accuracy of performance and time to completion. For all objectives, participants who are unable to complete the objective within a pre-specified time period are "pushed" to the next objective (referred to as "Forced Progression"). The primary VRFCAT endpoint is Total Adjusted Time to complete all 12 objectives. The assessment takes approximately 30 min to complete.

Key features of the VRFCAT are that it includes 6 alternate forms to minimize practice effects, eliminates administration and scoring errors through automatic delivery of stimuli and scoring, uses automatic cloud-based data transfer, and is regulatory compliant. A large normative database is available to compute standardized scores. In addition, the VRFCAT assesses contemporary skills that are well understood and performed across many cultural contexts to facilitate translatability for multi-national trials; it has been translated and culturally adapted into more than 40 languages.

Results from the original psychometric and validation demonstrated that the VRFCAT performance in schizophrenia patients demonstrates a large separation from healthy controls ( $d = 1.20$ ), good test-retest reliability ( $>0.80$ ), good distributional properties, minimal practice effects, an interpretable unidimensional structure.<sup>79,80</sup> Further, the VRFCAT demonstrated good convergent validity in terms of relations to traditional cognitive test battery performance, older measures of functional capacity (e.g., UPSA), and real-world functioning.<sup>79</sup> Strong convergent validity has subsequently been demonstrated in recent-onset patients and chronic institutionalized patients.<sup>81,82</sup> The VRFCAT has shown sensitivity to the effects of a computerized training interventions in schizophrenia and in older adults with Mild Cognitive Impairment.<sup>83,84</sup> The VRFCAT was migrated from a device resident computerized delivery system when it was developed to tablet-based system with direct cloud capture of performance data. A comparative study of the device resident and cloud/tablet-based delivery system found very high convergence between versions.<sup>54</sup> There was a slight, but systematic, performance advantage for tablet-based delivery because the tablet version does not rely on the ability to use a mouse.

Notably, a recent qualitative content validation study aimed at understanding consumer and other stakeholder perspectives on the real-world functional relevance VRFCAT<sup>85</sup> found convergent evidence that patients, caretakers, and treatment providers view the content of VRFCAT as highly relevant and important for achieving

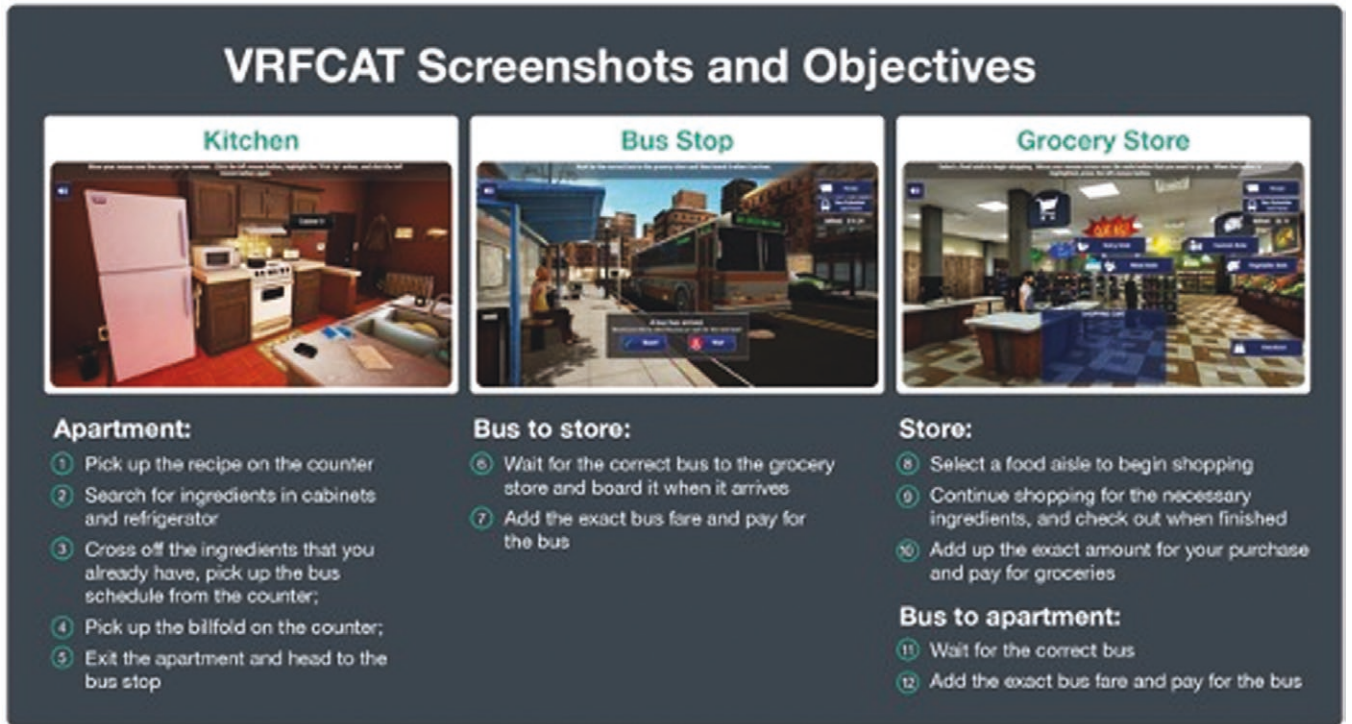


Fig. 3. Screenshots and 12 objectives from the virtual reality functional capacity assessment tool (VRFCAT).

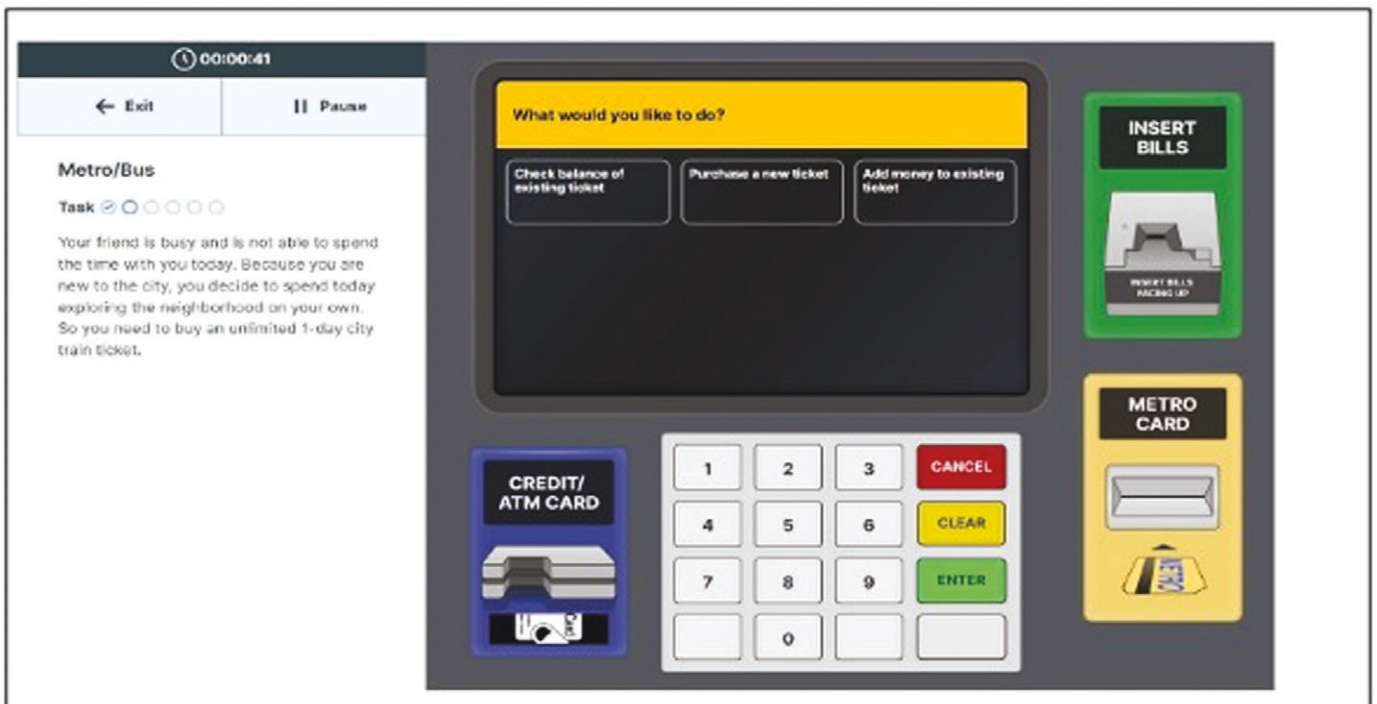


Fig. 4. Image of the ATM task from the functional skills assessment and training system (FUNSAT).

and maintaining functional independence. Based on the strong psychometric/quantitative and qualitative evidence, the VRFCAT has been accepted into FDAs COA Qualification Program to be considered for qualification

as a functional co-primary measure of functionally relevant changes in CIAS pharmacological clinical trials. The VRFCAT is already being used in several large international phase 2 and 3 clinical trials.



### *Functional Skills Assessment and Training System (FUNSAT)*

The FUNSAT<sup>86</sup> involves realistic modular simulations of contemporary technology-based functional skills. FUNSAT currently has 6 technology-related functional tasks: ATM and Internet Banking, Transit Ticket Purchase, Medication Label Comprehension and Organization, Telephone voice menu on a screen-based simulation of a mobile phone, and Pharmacy Website Utilization, including Refill, On-line shopping, and home delivery (see [figure 3](#)), with more tasks currently in development and validation. Five of the 6 simulations are completely like the real-world tasks, with touch screen delivery on displays that have the same size, layout, and visual characteristics as the real-world tasks, other than the presentation of instructions for elements of task performance that would be self-generated in real-world technology utilization. The outcomes examined across the simulations have included time to completion as well as errors, in addition to examining full mastery of tasks such that inferences regarding the real-world ability to successfully execute the skills can be made. Like the VRFCAT, for all objectives, participants who are unable to complete the objective within a pre-specified time period are “pushed” to the next objective (referred to as “Forced Progression”). Three alternative forms were developed and have been shown to be similar in difficulty and psychometric characteristics and to manifest minimal retest changes.

FUNSAT, like the VRFCAT, eliminates administration and scoring errors through automatic delivery of stimuli and scoring, uses automatic cloud-based data transfer, and is regulatory compliant. Like the VRFCAT, FUNSAT assesses contemporary skills that are well understood and performed across many cultural contexts to facilitate translatability for multi-national use. In the most recent treatment study, assessments were delivered in either English or Spanish and after an in-person orientation and baseline assessment, all training and follow-up assessments were delivered fully remotely and performed at home by participants. Software features allow for rapid transition and cultural adaptation to other languages and international variation in the task stimuli and a digital migration study similar to the VRFCAT showed high convergence between the original device resident software and the cloud-based touchscreen performance across all 6 tasks.<sup>54</sup>

The FUNSAT has been deployed across the same populations (i.e., schizophrenia and aging-related cognitive conditions) as the prior pencil-and-paper and prop-based functional capacity measures and has been shown to manifest correlations with cognitive performance consistent with measuring related but separable performance domains.<sup>87–89</sup> Healthy controls have been examined to potentially develop normative standards over time.

Cross-sectional correlations with legacy paper and pencil measures of functional capacity (UPSA) as well as other digital functional capacity measures (VRFCAT) have been demonstrated as well.

The FUNSAT has shown sensitivity to changes associated with remotely delivered interventions, including self-administered remote cognitive and functional skills training<sup>84,90</sup> (ES for FUNSAT Composite was  $d = 0.75$ ). Post-training performance in the FUNSAT has also been shown to correlate with changes in cognition, the VRFCAT, and real-world performance of the functional skills assessed by FUNSAT (assessed on a momentary basis with EMA). The FUNSAT has not been submitted for FDA regulatory clearance as an outcome measure in pharmacological trials, but the evidence regarding its reliability, validity, and sensitivity to both impairments in illness and response to treatments directly support its applicability ([figure 4](#)).

### *VR Grocery Shopping Tasks*

Shopping in a grocery store represents an important functional task for people with schizophrenia. Computer-based and VR shopping measures correlate with “real life” performance in grocery stores among those with schizophrenia.<sup>91</sup> Importantly, performance on VR grocery store tasks predicts variance in performance efficiency in actual grocery stores above and beyond traditional cognitive measures<sup>91</sup> in those with schizophrenia. Performance on VR grocery store tasks is also related to social functioning among those with schizophrenia in ways that traditional cognitive assessment tools are not.<sup>92</sup>

The earliest “virtual” grocery store assessment tasks were really computer tests, rather than true VR. For example, Laroi et al.<sup>92</sup> developed a computerized grocery shopping task in which participants were required to shop for a list of grocery store items using a joystick. By pressing buttons on the joystick, the participant was able to put items into a shopping cart, consult the shopping list, put items back on the shelf, move through the aisles, etc. Visual and auditory distractors were present, and various performance metrics were collected, including total time to complete the task, number of correct items, number of incorrect items, aisle redundancy, number of times the grocery list was consulted, etc. A group of 30 individuals with schizophrenia performed significantly worse than an equal number of healthy controls on most variables, with the total time to complete the task being the variable with the largest difference. A measure of social functioning (Global Assessment of Functioning) was correlated with performance on the shopping task but was not associated with performance on the traditional cognitive measures (except for a measure of cognitive inhibition), nor with the global cognitive score. This study reflected the early potential of VR-based grocery store measures for use in schizophrenia trials.



**Fig. 5.** Images from the VR cognition training system (VRCTS).

Another computer-based “virtual” grocery store task was developed by Greenberg et al.<sup>91</sup> whereby participants used a joystick to navigate and manipulate their environment to find the 10 items or pieces of information (e.g., the price of an item) required. Accuracy was measured by the total number of correct items selected. Efficiency was measured by the time taken to complete the task, and the number of aisles entered above the minimum required, when using the most efficient route. In a group of 43 individuals diagnosed with schizophrenia, performance on this task was correlated with “real life” performance in grocery stores.

Since these early studies, other more immersive VR grocery store assessment tools have been developed. Huang et al.<sup>93</sup> used an immersive VR system with a helmet called the VR Cognition Training System (VRCTS, see figure 5) in 32 Han Chinese individuals with schizophrenia and 25 healthy controls. The VRCTS consists of 2 tasks which entail finding common goods on a list of items found in a grocery store and putting them into a shopping cart using a joystick. There are 4 different levels with an increasing number of items on the list. Participants must memorize the list of items, but they are able to press a button on their joystick if they forget the list and it appears again. Scores are based on the correct number of items, errors, and completion time of each task. Huang et al. reported significant differences in completion time of the 2 shopping tasks between those with schizophrenia and healthy controls, and a significant correlation between performance on the VRCTS and the composite score of the MCCB.

### *Section Summary*

Though a range of technology-enhanced functional cognitive measures relevant to schizophrenia have been developed,<sup>53,94-97</sup> only a subset of them have been used in this disorder. The measures described above illustrate the range of real-world activities and technologies used thus far in schizophrenia. While all these measures are feasible to administer and have demonstrated separation of performance between individuals with schizophrenia and healthy controls, there are substantial differences in how extensively they have been validated and implemented in clinical trials. The more equipment and technology intensive tasks, like driving simulators and VR paradigms, have thus far been conducted in relatively small samples of people with schizophrenia by single academic research labs. The tablet-based VRFCAT and FUNSAT simulation tasks have considerably more validity evidence and documented psychometric properties, to support their use as endpoints in clinical trials. Both measures have already been successfully used in large multi-site, and even multi-country, trials in schizophrenia and other cognitively impaired populations.

### **Portable Device-Based Assessments in Schizophrenia**

Traditional in-person neuropsychological testing has long been the gold standard for cognitive assessment. Technological advances facilitate remote assessments of cognition and associated behaviors as participants engage in their daily activities in their natural environments. This section provides an overview of remotely delivered EMA cognitive and functional capacity measures, as

well complementary sensor-based data that is passively collected.

### *EMA Assessment of Cognition*

Unsupervised high-frequency cognitive EMA via smartphones is increasingly used to assess cognition remotely in clinical samples, including in-persons with cognitive impairment. Utilizing an EMA approach, cognitive EMA incorporates brief cognitive assessments, either “migrated” versions of traditional tasks or purpose-built mobile tasks, into daily life. This approach allows for the collection of multiple snapshots of cognitive performance throughout the day and/or across days, enabling a deeper understanding of the ongoing patterns of daily cognition.

One of the strengths of cognitive EMA is the ability to examine both average performance and variability in performance within and between individuals across time and contexts. This method is particularly valuable as cognition can fluctuate day-to-day and is greatly influenced by context, with emerging research suggesting that intra-individual variability (IIV) holds clinical relevance and may be a particularly sensitive marker of future cognitive decline.<sup>98–100</sup> Traditional neuropsychological testing makes it challenging to capture IIV over meaningful time frames and to identify other factors that may impact cognition both day-to-day and more long-term. Additionally, many traditional research methods focus on differences between individuals rather than capturing variations within individuals, an important gap that cognitive EMA can effectively address.

For example, in a sample of community-dwelling older adults, Schmitter-Edgecombe and colleagues<sup>101</sup> found that IIV on a tablet-based n-back test, administered 4 times daily for a week, showed a stronger association with self-reported functional status than an individual’s average performance on the task or their global neurocognitive functioning assessed in a laboratory setting. Within-day diurnal patterns of cognitive change have also been identified. In a sample of adults with type 1 diabetes, Hernandez and colleagues<sup>102</sup> administered a smartphone-based processing speed task 5 to 6 times per day for a 2-week period. They found a diurnal pattern such that processing speed was consistently slower in the morning and evening and fastest midday. This diurnal pattern of processing speed has been similarly identified in other populations, including normative data samples<sup>103</sup> as well as among people with Alzheimer’s disease and related dementias. In the context of schizophrenia, where long-term trajectories of cognitive impairment vary (some people improve over time, other remain unchanged, and some worsen, e.g.,<sup>104,105</sup>), cognitive EMA can offer a precision health approach to individually quantifying cognition in this population.

Smartphones, with their capabilities for interactive tasks, data recording, and real-time feedback, can administer cognitive tests across a wide range of cognitive functions in an ultra-brief manner, including memory, attention, executive functions, processing speed, language, and verbal skills, visuospatial ability, motor skills, and social cognition. Furthermore, mobile assessment allows for the integration of diverse contextual factors associated with cognition, such as self-reported mood, symptoms, functional behaviors, sleep, movement patterns, physical activity, and substance use.

EMA can also be used to collect complementary information about patients’ momentary perceptions of their performance on remotely administered cognitive tasks, or even of how well they see themselves performing actual daily life tasks. For example, brief EMA surveys have been used to collect appraisals of success/failure and effort/difficulty for social and goal-directed activities.<sup>106–108</sup> Unlike standard interview-based measures, which require individuals to retrospectively report their cognitive functioning over multiple weeks and often demonstrate poor convergent validity in schizophrenia,<sup>46,47</sup> EMA offers a promising alternative. EMA collects self-reported data about cognition in real-time, significantly reducing the demands on memory recall. As a result, momentary judgments of performance can be captured, which can then be compared to the global judgments collected from traditional self-report measures. This comparison may help quantify how challenges in momentary judgments contribute to response biases in global self-reports.

Cognitive EMA tests can be customized to the needs to various age groups, languages, populations, and research or clinical requirements. Adherence to intensive burst assessment in observational cognitive EMA studies in diverse patient populations, including people with schizophrenia, bipolar disorder, MCI/other dementia risk, and older persons with HIV, is consistently 70% or higher (e.g.,<sup>109–118</sup>). These bursts have varied in frequency and intensity, with some consisting of 1 test every other day for 30 days, while others involve 3–4 tests administered 4 times daily for 7 days, with various permutations in between.

Speech samples are another example of data that can be collected remotely, and the application of computational methods such as Natural Language Processing (NLP), which can automatically conduct speech analysis and extract speech features, holds promise for the detection of specific cognitive symptoms. Language-related alterations are hallmark features in people with schizophrenia, and include reduced word and speech production, use of simpler and shorter phrases, impaired processing of complex syntax, use of more word approximations, and repetitions, and reduced semantic verbal fluency. In a recent study, Silva et al.<sup>119</sup> found a decline in syntactic complexity in early psychosis, which was predictive of a diagnosis of schizophrenia. Additionally,

Liebenthal et al.<sup>120</sup> examined speech features and smartphone usage patterns in individuals with psychotic disorders and found that greater severity of conceptual disorganization symptoms was associated with increased speech verbosity, speech disfluency, missing smartphone data, and heightened smartphone usage during sleep time, suggesting digital measures of speech disfluency are scalable markers of clinical disorganization in this population. Integrating NLP with cognitive EMA would allow for the extraction of valuable insights from participants' test performance as well as natural speech, as it could ultimately help track changes in language use and cognitive patterns over time in response to treatment or lifestyle changes.

Although most studies have published psychometric properties of individual cognitive EMA tests, only 2 studies have reported correlations between cognitive EMA composites with lab-based neuropsychological testing composites, resulting in correlations ranging from  $r = 0.48$  to  $r = 0.53$ .<sup>110,113</sup> Furthermore, several clinical trials are underway utilizing cognitive EMA as a more sensitive outcome measure and to reduce the Number Needed to Treat. Unfortunately, most commercially available app-based tools for cognitive assessment lack validity data for their assessments.<sup>121</sup> To our knowledge, there are 11 app-based tools with published psychometric data in clinical samples<sup>113,118,122–131</sup>; we consider 2 examples with relatively well-developed evidence supporting their feasibility, acceptability, and validity in schizophrenia.

*NeuroUX*. NeuroUX has developed a wide suite of gamified cognitive EMA tests, many of which have been extensively validated in people with schizophrenia, as well as integration of these tests with passive sensor data (see “Functional Skills Assessment and Training System (FUNSAT)” Section) and EMA surveys. Currently, NeuroUX offers 17 different cognitive EMA tests assessing cognition in the following domains: Reaction time, recall and recognition memory, visuospatial memory, working memory, processing speed, attention, executive functions, and social cognition. Normative data in a sample of 394 English-speaking, U.S. residents (aged 20–79; 50% female; 70% White; 46% iOS users; 54% Android users) is currently available for eight of these tests (figure 6).<sup>132</sup> Each test takes ~1–2 min to complete. There are 2 tests from the NeuroUX battery that have the most extensive psychometric data in people with schizophrenia: the Mobile Variability Difficulty List Memory Tests (VLMT) and the EMA facial recognition test, which are described in more detail below.

The VLMT is comprised of 3 different list lengths (6-, 12-, and 18- words) and the development of this task has been previously described.<sup>111,115</sup> In a study evaluating the validity of repeated administration of the VLMT among people with schizophrenia-spectrum disorders ( $n = 98$ ) and bipolar disorder ( $n = 70$ ), performance on the VLMT

12- and 18-word lists was positively correlated with the a laboratory assessment of the Hopkins Verbal Learning Test ( $\rho = 0.52$ ,  $P < .001$ ) and the UPSA ( $\rho = 0.41$ ,  $P < .001$ ).<sup>111</sup> Adherence to the cognitive EMA protocol for the whole sample was 75.3%, and adherence did not differ by diagnostic status or correlate with cognitive or symptom variables, indicating that the task can be used by a wide constituency. No practice effects were found for the alternative forms of the 6-, 12- or 18-item VLMT lists. As expected, performance parametrically declined with longer word lists, and people with schizophrenia performed worse on the VLMT than people with bipolar disorder.

In addition to the ability to assess traditional neuropsychological domains (memory, executive functions, processing speed, etc.), NeuroUX developed a cognitive EMA test of social cognition designed for use in people with serious mental illness. In a validation study of the EMA facial recognition test, people with psychotic disorders ( $n = 86$ ; nonaffective and affective psychosis) completed this task on their smartphones 3 times daily for a 10-day period. Convergent validity was found between performance on the mobile tasks and in-lab measures of affect recognition.<sup>112</sup> Further, better performance on the cognitive EMA facial affect recognition test was associated with greater pleasure and more positive appraisals of others during social interactions.<sup>133</sup> A recently published study by Parrish et al.<sup>134</sup> used this task to examine relationships between suicide risk factors and facial affect recognition in a sample of 273 people with serious mental illness. This study found that real-time perceptions of threat (measured via EMA) are related to constructs relevant to suicidal ideation in this population, more-so than general facial affect recognition, which holds implications for suicide prevention work.

Bridging the gap between cognitive assessment and functional outcomes, a randomized clinical trial is currently underway (clinicaltrials.gov ID: NCT05899348) of a newly developed smartphone-based cognitive training program targeting introspective accuracy, or the second-order judgements of one's cognitive and functional abilities as compared to objective evidence, as a mechanism by which to improve functional outcomes in people with psychotic disorders. This program, called *Improving Thinking through Everyday Self-Assessment Training* (iTEST), utilizes the VLMT and EMA facial recognition tests to conduct task-based trainings of introspective accuracy and is assesses near and far transfer of training to daily life.

*mindLAMP*. The mindLAMP app, an open-source mental health tool, offers EMA surveys, mobile cognitive tests, and access to capturing passive smartphone data streams.<sup>135</sup>

Validation studies have been published with two of the mindLAMP mobile cognitive tests in people with



**Fig. 6.** Screenshots of NeuroUX's mobile cognitive tests: (A) memory list; (B) memory matrix; (C) matching pair; (D) quick tap 1; (E) quick tap 2; (F) odd one out; (G) CopyKat; (H) hand swype.

schizophrenia: the Jewels Trail Tests A and B (smartphone versions of the Trail Making Test). These studies demonstrated the reliability and validity of these tests compared to lab-based Trail Making Tests, in the U.S. and in India, although sample sizes were relatively small (ranging from 18 to 76 people with schizophrenia).<sup>136,137</sup> Another study investigated the associations between screen time and real-world cognition, as measured by the Jewels Trail Test B, among individuals with schizophrenia and found while screen time did not correlate with lab-based cognitive assessment (using the BACS), it was linked to longitudinal beta values for the Jewels B cognitive test on the mindLAMP app ( $F = 5.43$ ,  $P < .001$ ). Notably, this association was observed only among people with schizophrenia, as there was no significant relationships between Jewels B performance and screen time among healthy controls.<sup>138</sup>

Moving beyond small, single lab studies, mindLAMP's EMA and digital phenotyping platform is currently being used as an assessment tool in the Accelerating Medicines Partnership (AMP) Schizophrenia program. This global initiative aims to develop tools that will expedite the

development of effective early-stage treatments for individuals at risk for schizophrenia.

#### *Integrating Passive Sensor Data*

Passive measures, including smartphone sensors and wearable devices, can be valuable tools in research on cognition in schizophrenia, offering unique insights into individuals' daily lives and behaviors. Examples of data related to cognition and functioning that can be captured passively, with no effort or even awareness on the part of subjects, include movement patterns, sleep quality, social interactions, and environmental factors. Given that people with schizophrenia often exhibit deficits in their insight or impairments in recall, passive measures can provide objective and continuous assessment of their real-world functioning. Integrating passive measures with cognitive EMA can provide additional contextual information and corroborate performance data. The pairing of active and passive data can enhance the richness of data collected and well as offer a more comprehensive understanding of the dynamic interplay between

symptoms, behavior, cognition, and functioning, and holds promise for informing personalized interventions.

Global Positioning System (GPS) data represents 1 passive data stream that has demonstrated validity as a metric of functioning in people with schizophrenia. For instance, the severity of avolition can be characterized by combining EMA surveys and GPS coordinates to determine if an individual spends most of their time home, alone, and engaged in unproductive activities such as pacing, smoking, watching TV, resting, sitting alone, or doing nothing. Numerous investigations have found strong convergence between self-reported (EMA) and GPS-based location measures in individuals diagnosed with schizophrenia, with both data streams significantly correlating with clinician ratings of negative symptoms (e.g.,<sup>139,140</sup>; see Daniel et al., 2023 for a comprehensive review of remote assessment of negative symptoms of schizophrenia<sup>141</sup>).

Another source of passively collected data is meta-data is derived from interactions with mobile devices. This can include number of text or e-mail messages sent, information searches, and typing speed. These indicators are also easily obtained and can be used to predict other elements of momentary behaviors, including sustaining stable daily routines<sup>142</sup> or larger issues such as impending relapse.<sup>143</sup> The interaction between different elements of passively measured information streams can also be monitored remotely and provides the possibility of customized interventions based on continuous assessment of passively collected data.

Overall, the integration of passive measures with cognitive EMA holds significant promise in advancing our understanding of cognitive processes in schizophrenia and could provide valuable insights into real-world manifestations of symptoms and functioning in people with schizophrenia. Both NeuroUX and AMP-SZ include access to collection of passive data streams in their platforms, to be utilized in conjunction with the mobile cognitive tasks or independently, thus streamlining the data collection process. To date, few studies have used digital phenotyping approaches in people with schizophrenia (see Lane et al. for a review<sup>144,145</sup>), but there is promise that this method could enable a more holistic approach to assessing cognitive functioning, as well as enhance our ability the link cognitive abilities and daily functioning.

#### *Other Remote Assessment Applications in Schizophrenia*

The lessons learned in the COVID pandemic have found that some interview and performance-based functional capacity measures can be assessed on a remote basis. For informant based functional capacity interviews, such as the SCoRS, the CAI, or interactive verbal-only strategies (such as the SSPA), video-conferencing strategies have

been shown to be easily adaptable. Both for participant and informant data collection, there is no requirement for manipulation of props or stimulus items, as are required for paper and pencil neuropsychological tests or functional capacity assessments. In fact, recent published studies have used data collected in-person or with video conferencing for both the CAI and SSPA, with no differences in performance associated with data collection modality.<sup>146</sup>

Regarding performance-based VRFCAT and FUNSAT functional capacity used in schizophrenia and other cognitively impaired populations, the VRFCAT has been delivered exclusively on an in-person basis. Even in studies with remote training, the VRFCAT assessment was performed in-person.<sup>83</sup> In contrast, the current version of FUNSAT is cloud-based and fully remotely deliverable. In a recent validation study of the FUNSAT software in MCI and cognitively unimpaired individuals, remotely captured performance on the FUNSAT was found to be feasible, with over 90% adherence and valid data collected in MCI and cognitively unimpaired population.<sup>84</sup> Further, both remotely delivered FUNSAT assessments and in-person VRFCAT assessments were sensitive to the effects of a remotely-delivered combined CCT and skills training intervention and were found to improve by very similar amounts from baseline to the end of training (ES for FUNSAT Composite was  $d = 0.75$ ; ES for VRFCAT total adjusted time to completion:  $d = 0.64$ ).<sup>84,90,147</sup> The FUNSAT system with remote delivery has been used as an outcome measure primarily in trials of MCI. Large-scale clinical initiatives in SMI populations are underway and will provide similar feasibility data for remote assessments in this population.

#### *Section Summary*

Collectively, these findings demonstrate remotely deliverable cognitive EMA and functional capacity measures, and even fully remote cognitive and skills training interventions, have proven feasible and acceptable in participants with schizophrenia and in other cognitively impaired populations. Use of cognitive EMA may allow for a deeper understanding of daily cognitive functioning as well as intra-individual variability over time. While most commercially available app-based tools for cognitive assessment lack validity data, 2 exceptions are the NeuroUX battery and mindLAMP, which have both produced initial validity data in individuals with schizophrenia. Remote collection of complementary passive data, such as GPS, and voice, also demonstrates initial validity in those with schizophrenia. In addition, functional capacity interviews are amenable to administration outside of traditional research settings via video-conferencing technology in schizophrenia, and initial studies support the validity of remote, unsupervised

administration of the performance-based FUNSAT in people with MCI.

## Conclusions

Cognitive and functional capacity assessment developers have embraced technological advances in the early 21st century. This has led to a surge of innovative measures and methods that show considerable promise for enhancing the ecological validity, including both verisimilitude and veridicality, of assessments conducted both inside and outside of a clinical research setting. There is solid evidence for the feasibility and initial validity of all the types of measures considered in schizophrenia. Furthermore, development and validation of the VRFCAT, FUNSAT, and some remote cognitive EMA measures are already at a relatively advanced stage. Driving simulation has also been used extensively in clinical trials, though not specifically in schizophrenia. The hope is that more ecologically valid cognitive measures can be used in observational studies to help us identify more functionally impactful targets for psychosocial and pharmacological interventions in schizophrenia. Ideally, these tasks could also serve as sensitive endpoints in clinical trials to evaluate the efficacy of novel interventions aimed at improving functional outcomes.

## Remaining Challenges

The proliferation of technology-enhanced measures reflects an exciting new chapter in the history of cognitive and functional outcome assessments in schizophrenia. Whether these measure will ultimately address key limitations of traditional neuropsychological tasks remains to be seen. In the context of rapidly emerging and alluringly marketed new technologies, validation needs to catch up with development. For the promise of new ecological measures to be fully realized, the field will need to address practical, scientific, and regulatory challenges.

Practical challenges include factors such as accessibility, compatibility, costs, staff training needs, and participant support needs for new technology-enhanced assessments. For example, although VR paradigms offer the potential for in-clinic assessments with unprecedented verisimilitude, their development and uptake have been hampered by expensive specialized equipment and resource requirements. Major advances in terms of verisimilitude, compactness, hardware-software integration, and reduced costs have increased the practicality of conducting larger scale VR validation studies in recent years.<sup>148</sup> However, further method development is clearly needed for scalable implementation of VR assessments in large multi-site trials.

For unsupervised cognitive EMA and passive remote measures, additional practical challenges include attributability, dealing with nonadherence and missing

data, inter-operability between different hardware and operating systems, and ensuring data security and privacy.<sup>149</sup> Notably, due to privacy concerns, some manufacturers and regional legislative bodies are making it more difficult to collect certain types of data (e.g., GPS) from smartphones. Further, although remote digital measures have the potential to substantially expand geographic and participant diversity representation in clinical trials, it is critical to consider potential access and acceptance barriers associated with factors such as age, socioeconomic status, and cultural characteristics.

From a scientific perspective, it is critical for researchers to be clear about how they operationalize “ecological validity” when developing technology-enhanced methods. Recent reviews document the diverse types of evidence that have been used to support claims of a test’s ecological validity (e.g., predict daily functioning, differentiated clinical groups, correlated with other cognitive tests, has face validity).<sup>150-152</sup> Consequently, the term has been criticized for being conceptually vague and potentially misleading, emphasizing the need for researchers to specify the functional context of the cognitive processes in which they are interested.<sup>150</sup> In schizophrenia, the goal of using ecologically valid tests is typically to more precisely identify the determinants of daily life functional difficulties (i.e., veridicality together with verisimilitude) so we can develop and test interventions that more effectively target them. For technology-enhanced functional capacity measures used in clinical research settings, the scientific requirements are largely the same as for any new performance-based outcome assessment: strong quantitative evidence for their psychometric characteristics and convergent/discriminant validity, as well as qualitative evidence supporting their content validity, is required for acceptance by regulators.<sup>153,154</sup> Along these lines, test developers and regulators will need to agree on what constitutes a “ground truth” daily life functioning metric to establish their validity for use in clinical trials.

However, efforts to validate new technology-enhanced methods raise some unique complications, particularly for remotely collected assessments. For unsupervised cognitive EMA and passive remote measures, all the standard psychometric and validation requirements for in-clinic assessments are applicable for regulatory acceptance.<sup>149</sup> Although legacy cognitive and clinical measures created decades ago are often considered gold standards for validation from a regulatory perspective, it is possible that novel remote measures can provide superior precision or even capture qualitatively distinct aspects of functioning, resulting in relatively lower convergence with traditional measures as compared to the typical standards applied to evaluating convergent validity among traditional clinic-based measures. What is an appropriate level of convergence with traditional gold standard clinical measures? In addition, passive sensor-based measures collect very large quantities of data that

are typically processed using (often proprietary) artificial intelligence-based algorithms to create clinically meaningful outcome variables. These AI-derived metrics will need to be sufficiently transparent and understandable for regulatory acceptance.<sup>155,156</sup>

### Looking Forward

Looking forward, we expect the current proliferation of technology-enhanced ecologically valid assessments is unlikely to slow down. We anticipate expansion of clinic-based VR approaches to other functionally important areas, such as using chatbots and avatars, which are increasingly sophisticated, to provide veridical assessments of social cognitive processes. The trend toward increasingly hybrid in-clinic/remote assessment approaches is likely to continue; in fact, fully remote trials that provide both computerized cognitive remediation or functional skills training and clinical assessments almost entirely outside of a conventional clinical research setting are already happening.<sup>84,90,147</sup> Finally, despite the challenges associated with remote passive metrics, we expect increased development efforts in this realm due to the appeal of identifying objective indices of real-world cognition and functioning that involve no deliberate effort aside from wearing or carrying a digital device.

### Financial Disclosures

Dr Horan is an employee Bristol Myers Squibb. Dr Moore is a cofounder of KeyWise AI (developer of passively collected smartphone keystroke digital biomarkers) and NeuroUX (developer of a smartphone platform that delivers EMA, mobile cognitive tests, and ability to collect passive sensing metrics). Dr Harvey is a consultant for Alkermes, Boehringer-Ingelheim, Karuna Therapeutics, Merck, Minerva Neuroscience, and Sunovion/DSP and has received royalties from WCG. He is also chief scientific officer of i-Function, which created and distributes FUNSAT.

### References

1. Reitan RM, Wolfson D. *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*. Vol 4. Neuropsychology Press; 1985.
2. Green MF, Horan WP, Lee J. Nonsocial and social cognition in schizophrenia: current evidence and future directions. *World Psychiatry*. 2019;18(2):146–161.
3. McCutcheon RA, Keefe RSE, McGuire PK. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. *Mol Psychiatry*. 2023;28:1902–1918.
4. Kahn RS, Keefe RS. Schizophrenia is a cognitive illness: time for a change in focus. *JAMA Psychiatry*. 2013;70(10):1107–1112.
5. Green MF, Nuechterlein KH. Should schizophrenia be treated as a neurocognitive disorder? *Schizophr Bull*. 1999;25(2):309–319.
6. Nuechterlein KH, Green MF. *MATRICES Consensus Cognitive Battery*. Los Angeles: MATRICS Assessment, Inc.; 2006.
7. Nuechterlein KH, Green MF, Kern RS. The MATRICS consensus cognitive battery: an update. *Curr Top Behav Neurosci*. 2023;63:1–18.
8. Horan WP, Catalano LT, Green MF. An update on treatment of cognitive impairment associated with schizophrenia. *Curr Top Behav Neurosci*. 2023;63:407–436.
9. Leonberger FT. The question of organicity: is it still functional? *Prof Psychol Res Pract*. 1989;20(6):411–414.
10. Lewis DA, Sweet RA. Schizophrenia from a neural circuitry perspective: advancing toward rational pharmacological therapies. *J Clin Invest*. 2009;119(4):706–716.
11. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychol Rev*. 2003;13(4):181–197.
12. Weizenbaum E, Torous J, Fulford D. Cognition in context: understanding the everyday predictors of cognitive performance in a new era of measurement. *JMIR Mhealth Uhealth* 2020;8(7):e14328.
13. Franzen MD, Wilhelm KL. Conceptual foundations of ecological validity in neuropsychology. In: Sbordone RJ, Long CJ, eds. *Ecological Validity of Neuropsychological Testing*. Gr Press/St Lucie Press, Inc.; 1996:91–112.
14. Sbordone RJ. Ecological validity: some critical issues for the neuropsychologist. In: Sbordone RJ, Long CJ, eds. *Ecological Validity of Neuropsychological Testing*. Gr Press/St Lucie Press, Inc.; 1996:15–41.
15. Burgess PW, Alderman N, Forbes C, et al. The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *J Int Neuropsychol Soc*. 2006;12(2):194–209.
16. Othuba RA, van den Broek MD, Johns LC. Ecological validity of measures of executive functioning. *Br J Clin Psychol*. 2005;44(Pt 2):269–278.
17. Spooner DM, Pachana NA. Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. *Arch Clin Neuropsychol*. 2006;21(4):327–337.
18. Tyburski E, Mak M, Sokolowski A, et al. Executive dysfunctions in schizophrenia: a critical review of traditional, ecological, and virtual reality assessments. *J Clin Med* 2021;10(13):2782.
19. RK H, GJ C, JL T, GG K, G C. *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Odessa, FL: Psychological Assessment Resources; 1993.
20. Wechsler D. *Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)*. San Antonio, TX: The Psychological Corporation; 2008.
21. Keefe RSE. Assessment of cognition in schizophrenia treatment studies. In: Harvey PD, ed. *Cognitive Impairment in Schizophrenia*. Cambridge University Press; 2013:231–246.
22. Barnett JH, Robbins TW, Leeson VC, Sahakian BJ, Joyce EM, Blackwell AD. Assessing cognitive function in clinical trials of schizophrenia. *Neurosci Biobehav Rev*. 2010;34(8):1161–1177.
23. Koren D, Seidman LJ, Goldsmith M, Harvey PD. Real-world cognitive--and metacognitive--dysfunction in schizophrenia: a new approach for measuring (and remediating) more “right stuff.”. *Schizophr Bull*. 2006;32(2):310–326.
24. McGurk SR, Twamley EW, Sitzer DI, McHugo GJ, Mueser KT. A meta-analysis of cognitive remediation in schizophrenia. *Am J Psychiatry*. 2007;164(12):1791–1802.



25. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the “right stuff?”. *Schizophr Bull.* 2000;26(1):119–136.
26. Green MF, Llerena K, Kern RS. The “Right Stuff” Revisited: what have we learned about the determinants of daily functioning in schizophrenia? *Schizophr Bull.* 2015;41(4):781–785.
27. Kerr SL, Neale JM. Emotion perception in schizophrenia: specific deficit or further evidence of generalized poor performance? *J Abnorm Psychol.* 1993;102(2):312–318.
28. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the mind in the eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry Allied Discip.* 2001;42:241–251.
29. Robertson I, Ward T, Ridgeway V, Nimmo-Smith I. *The Test of Everyday Attention.* London, England: Pearson Assessment.; 1994.
30. Wilson BA, Alderman N, Burgess PW, Emslie H, Evans JJ. *Behavioural Assessment of the Dysexecutive Syndrome.* Bury St Edmunds, UK: Harcourt Assessment.; 1996.
31. Fett AK, Viechtbauer W, Dominguez MD, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neurosci Biobehav Rev.* 2011;35:573–588.
32. Penn DL, Corrigan PW, Bentall RP, Racenstein JM, Newman L. Social cognition in schizophrenia. *Psychol Bull.* 1997;121:114–132.
33. Green MF, Olivier B, Crawley JN, Penn DL, Silverstein S. Social cognition in schizophrenia: recommendations from the MATRICS New Approaches Conference. *Schizophr Bull.* 2005;31:882–887.
34. Pinkham AE, Harvey PD, Penn DL. Social cognition psychometric evaluation: results of the Final Validation Study. *Schizophr Bull.* 2018;44(4):737–748.
35. Nijman SA, Veling W, van der Stouwe ECD, Pijnenborg GHM. Social cognition training for people with a psychotic disorder: a network meta-analysis. *Schizophr Bull.* 2020;46(5):1086–1103.
36. Davis MC, Green MF, Lee J, et al. Oxytocin-augmented social cognitive skills training in schizophrenia. *Neuropsychopharmacology.* 2014;39(9):2070–2077.
37. Horan WP, Green MF. Treatment of social cognition in schizophrenia: current status and future directions. *Schizophr Res.* 2019;203:3–11.
38. Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. *Brain.* 1991;114 (Pt 2)(Pt 2):727–741.
39. Bellack AS, Brown CH, Thomas-Lohrman S. Psychometric characteristics of role-play assessments of social skill in schizophrenia. *Behav Ther.* 2006;37(4):339–352.
40. Baum CM, Wolf TJ, Wong AWK, et al. Validation and clinical utility of the executive function performance test in persons with traumatic brain injury. *Neuropsychol Rehabil.* 2017;27(5):603–617.
41. Harvey PD, Velligan DI, Bellack AS. Performance-based measures of functional skills: usefulness in clinical treatment studies. *Schizophr Bull.* 2007;33(5):1138–1148.
42. Harvey PD, Horan WP, Atkins AS, et al. Factor structure of cognitive performance and functional capacity in schizophrenia: evidence for differences across functional capacity measures. *Schizophr Res.* 2020;223:297–304.
43. Buchanan RW, Davis M, Goff D, et al. A summary of the FDA-NIMH-MATRICES workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophr Bull.* 2005;31(1):5–19.
44. Keefe RSE, Poe M, Walker TM, Kang JW, Harvey PD. The schizophrenia cognition rating scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *Am J Psychiatry.* 2006;163:426–432.
45. Ventura J, Reise SP, Keefe RS, et al. The Cognitive Assessment Interview (CAI): development and validation of an empirically derived, brief interview-based measure of cognition. *Schizophr Res.* 2010;121(1-3):24–31.
46. Morgan O, Strassnig MT, Moore RC, et al. Accuracy of immediate self-assessment of neurocognitive test performance: associations with psychiatric diagnosis and longitudinal psychotic symptoms. *J Psychiatr Res.* 2022;156:594–601.
47. Silberstein J, Harvey PD. Impaired introspective accuracy in schizophrenia: an independent predictor of functional outcomes. *Cogn Neuropsychiatry.* 2019;24(1):28–39.
48. Green MF, Nuechterlein KH, Kern RS, et al. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry.* 2008;165(2):221–228.
49. Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV. UCSD Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull.* 2001;27(2):235–245.
50. Bellack AS, Sayers M, Mueser KT, Bennett M. Evaluation of social problem solving in schizophrenia. *J Abnorm Psychol.* 1994;103(2):371–378.
51. Bowie CR, Twamley EW, Anderson H, Halpern B, Patterson TL, Harvey PD. Self-assessment of functional status in schizophrenia. *J Psychiatr Res.* 2007;41(12):1012–1018.
52. Keefe RS, Poe M, Walker TM, Harvey PD. The relationship of the brief assessment of cognition in schizophrenia (BACS) to functional capacity and real-world functional outcome. *J Clin Exp Neuropsychol.* 2006;28(2):260–269.
53. Pieri L, Tosi G, Romano D. Virtual reality technology in neuropsychological testing: a systematic review. *J Neuropsychol.* 2023;17(2):382–399.
54. Harvey PD, Keefe RSE, Kallestrup P, Czaja SJ, Klein H, Horan WP. Migration of digital functional capacity assessments from device resident to cloud-based delivery: development and convergent validity. *Schizophr Res: Cognit.* in press.
55. Palmer BW, Heaton RK, Gladsjo JA, et al. Heterogeneity in functional status among older outpatients with schizophrenia: employment history, living situation, and driving. *Schizophr Res.* 2002;55(3):205–215.
56. Food and Drug Administration (FDA). Evaluating drug effects on the ability to operate a motor vehicle: guidance for industry. In: Services USDoHaH, ed. Silver Spring, MD: Center for Drug Evaluation and Research (CDER); 2017:1–9.
57. Caffo AO, Tinella L, Lopez A, et al. The drives for driving simulation: a scientometric analysis and a selective review of reviews on simulated driving research. *Front Psychol.* 2020;11:917.
58. Wynne RA, Beanland V, Salmon PM. Systematic review of driving simulator validation studies. *Saf Sci.* 2019;117:138–151.
59. Brown T, Dow B, Marshall D, Allen S. Validation of stopping and turning behavior for novice drivers in the National Advanced Driving Simulator. Available at: <https://www.nads-sc.uiowa.edu/dscna/2007/papers/Section%203B%20-%20Simulator%20Characteristics%20-%20Applications%20II/Brown.pdf>

60. Salaani MK, Heydinger GJ, Grygier PA. Measurement and modeling of tire forces on low coefficient surface (2006-01-0559). *SAE International*. Available at: [https://www.nhtsa.gov/sites/nhtsa.gov/files/nadssa\\_e\\_paper2006010559.pdf](https://www.nhtsa.gov/sites/nhtsa.gov/files/nadssa_e_paper2006010559.pdf)
61. Lee JD, Ward N, Boer E, Brown TL, Balk SA, Ahmad O. Exploratory advanced research: Making driving simulators more useful for behavioral research—simulator characteristics comparison and model-based transformation. Available at: [http://www.nads-sc.uiowa.edu/publicationStorage/20131399331159.N2013-016\\_Making%20driving%20simul.pdf](http://www.nads-sc.uiowa.edu/publicationStorage/20131399331159.N2013-016_Making%20driving%20simul.pdf). Accessed 5 January, 2016.
62. Senserrick TM, Brown T, Quistberg DA, Marshall D, Winston FK. Validation of simulated assessment of teen driver speed management on rural roads. *Annu Proc Assoc Adv Automot Med*. 2007;51:525–536.
63. Kay GG, Hochadel T, Sicard E, Natarajan KK, Kim NN. Next-day residual effects of flibanserin on simulated driving performance in premenopausal women. *Hum Psychopharmacol*. 2017;32(4):e2603.
64. Kay GG, Schwartz HI, Wingertzahn MA, Jayawardena S, Rosenberg RP. Next-day residual effects of gabapentin, diphenhydramine, and triazolam on simulated driving performance in healthy volunteers: a phase 3, randomized, double-blind, placebo-controlled, crossover trial. *Hum Psychopharmacol*. 2016;31(3):217–226.
65. Kay GG, Berman B, Mockoviak SH, et al. Initial and steady-state effects of diphenhydramine and loratadine on sedation, cognition, mood, and psychomotor performance. *Arch Intern Med*. 1997;157(20):2350–2356.
66. Su S, Kay G, Hochadel T, et al. A randomized, multicenter trial assessing the effects of rapastinel compared to ketamine, alprazolam, and placebo on simulated driving performance. *Clin Transl Sci*. 2022;15(1):255–266.
67. Verster JC, van de Loo AJ, Roth T. Mirtazapine as positive control drug in studies examining the effects of antidepressants on driving ability. *Eur J Pharmacol*. 2015;753:252–256.
68. Muehlan C, Brooks S, Vaillant C, et al. Driving performance after bedtime administration of daridorexant, assessed in a sensitive simulator. *Clin Pharmacol Ther*. 2022;111(6):1334–1342.
69. St. Germain S, Kurtz M. Virtual driving in individuals with schizophrenia. *Annual Review Cyber Therapy Telemedicine*. 2004;2:153–159.
70. Wylie KR, Thompson DJ, Wildgust HJ. Effects of depot neuroleptics on driving performance in chronic schizophrenic patients. *J Neurol Neurosurg Psychiatry*. 1993;56(8):910–913.
71. St Germain SA, Kurtz MM, Pearlson GD, Astur RS. Driving simulator performance in schizophrenia. *Schizophr Res*. 2005;74:121–122.
72. Iwata M, Iwamoto K, Kawano N, Kawaue T, Ozaki N. Evaluation method regarding the effect of psychotropic drugs on driving performance: a literature review. *Psychiatry Clin Neurosci*. 2018;72(10):747–773.
73. Okada H, Sawamura D, Kunita K, et al. Prefrontal activation during simulated driving in people with schizophrenia: A functional near-infrared spectroscopy study. *Psychiatry Res*. 2023;326:115285.
74. Fuermaier ABM, Piersma D, Huntjens RJC, et al. Simulated car driving and its association with cognitive abilities in patients with schizophrenia. *Schizophr Res*. 2019;204:171–177.
75. Brunnauer A, Laux G, Zwick S. Driving simulator performance and psychomotor functions of schizophrenic patients treated with antipsychotics. *Eur Arch Psychiatry Clin Neurosci*. 2009;259(8):483–489.
76. Kagerer S, Winter C, Moller HJ, M S. Effects of haloperidol and atypical neuroleptics on psychomotor performance and driving ability in schizophrenic patients. *Neuropsychobiology*. 2003;47:212–218.
77. Soyka M, Winter C, Kagerer S, Brunnauer M, Laux G, Moller HJ. Effects of haloperidol and risperidone on psychomotor performance relevant to driving ability in schizophrenic patients compared to healthy controls. *J Psychiatr Res*. 2005;39(1):101–108.
78. Biedermann F, Kurzhäler I, Haibach M, et al. Driving fitness in clinically stable outpatients with chronic schizophrenia. *Compr Psychiatry*. 2022;118:152340.
79. Keefe RSE, Davis VG, Atkins AS, et al. Validation of a Computerized test of Functional Capacity. *Schizophr Res*. 2016;175(1-3):90–96.
80. Harvey PD, Khan A, Atkins A, Keefe RS. Virtual reality assessment of functional capacity in people with schizophrenia: associations with reduced emotional experience and prediction of functional outcomes. *Psychiatry Res*. 2019;277:58–63.
81. Ventura J, Welikson T, Ered A, et al. Virtual reality assessment of functional capacity in the early course of schizophrenia: associations with cognitive performance and daily functioning. *Early Interv Psychiatry*. 2020;14(1):106–114.
82. Lindenmayer JP, Goldring A, Borne S, et al. Assessing instrumental activities of daily living (iADL) with a game-based assessment for individuals with schizophrenia. *Schizophr Res*. 2020;223:166–172.
83. Nahum M, Lee H, Fisher M, et al. Online social cognition training in schizophrenia: a double-blind, randomized, controlled multi-site clinical trial. *Schizophr Bull*. 2021;47(1):108–117.
84. Harvey PD, Chirino M, Mueller A, et al. Improvements in performance based measures of functional capacity and cognition after computerized functional skills training in older people with mild cognitive impairment and healthy comparators. *Psychiatry Res*. 2024;334:115792.
85. Horan WP, Depp CA, Hurst S, et al. Qualitative analysis of the content validity of the virtual reality functional capacity assessment tool (VRFCAT) in Schizophrenia: a multi-stakeholder perspective. *Schizophr. Bull. Open*. 2023;4(1):sgad012.
86. Harvey PD, Tibirić L, Kallestrup P, Czaja SJ. A computerized functional skills assessment and training program targeting technology based everyday functional skills. *J Vis Exp*. 2020;(156):e603300.
87. Harvey PD, Forero DB, Ahern LB, Tibirica L, Kallestrup P, Czaja SJ. The Computerized functional skills assessment and training program: sensitivity to global cognitive impairment, correlations with cognitive abilities, and factor structure. *Am J Geriatr Psychiatry*. 2021;29(4):395–404.
88. Czaja SJ, Loewenstein DA, Lee CC, Fu SH, Harvey PD. Assessing functional performance using computer-based simulations of everyday activities. *Schizophr Res*. 2017;183:130–136.
89. Czaja SJ, Loewenstein DA, Sabbag SA, Curiel RE, Crocco E, Harvey PD. A novel method for direct assessment of everyday competence among older adults. *J Alzheimers Dis*. 2017;57(4):1229–1238.
90. Dowell-Esquivel C, Czaja SJ, Kallestrup P, Depp CA, Saber JN, Harvey PD. Computerized cognitive and skills training in older people with mild cognitive impairment: using ecological momentary assessment to index treatment-related changes in

- real-world performance of technology-dependent functional tasks. *Am J Geriatr Psychiatry*. 2023;32:446–459.
91. Greenwood KE, Morris R, Smith V, Jones AM, Pearman D, Wykes T. Virtual shopping: A viable alternative to direct assessment of real life function? *Schizophr Res*. 2016;172(1-3):206–210.
  92. Laroi F, Canlaire J, Mourad H, M VDL. Relations between a computerized shopping task and cognitive tests in a group of persons diagnosed with schizophrenia compared with healthy controls. *J Int Neuropsychol Soc*. 2010;16(1):180–189.
  93. Huang B, Li S, Sun B, et al. Verification of using virtual reality to evaluate deficiencies in cognitive function among patients with schizophrenia in the remission stage: a cross-sectional study. *BMC Psychiatry*. 2021;21(1):21.
  94. Hsu WY, Goverover Y, Bove RM. Capturing cognitive changes in multiple sclerosis by performance-based functional and virtual reality assessments. *Ann Phys Rehabil Med*. 2023;66(3):101677.
  95. Kirkham R, Kooijman L, Albertella L, Myles D, Yucel M, Rotaru K. Immersive virtual reality-based methods for assessing executive functioning: systematic review. *JMIR Serious Games*. 2024;12:e50282.
  96. Parsons TD, Gaggioli A, Riva G. Extended reality for the clinical, affective, and social neurosciences. *Brain Sci*. 2020;10(12):922.
  97. Reggente N. VR for Cognition and Memory. *Curr Top Behav Neurosci*. 2023;65:189–232.
  98. Holmqvist SL, Thomas KR, Brenner EK, et al. Longitudinal intraindividual cognitive variability is associated with reduction in regional cerebral blood flow among Alzheimer's Disease biomarker-positive older adults. *Front Aging Neurosci*. 2022;14:859873.
  99. Lövdén M, Li SC, Shing YL, Lindenberger U. Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: longitudinal data from the Berlin Aging Study. *Neuropsychologia*. 2007;45(12):2827–2838.
  100. MacDonald SW, Li SC, Bäckman L. Neural underpinnings of within-person variability in cognitive functioning. *Psychol Aging*. 2009;24(4):792–808.
  101. Schmitter-Edgecombe M, Sumida C, Cook DJ. Bridging the gap between performance-based assessment and self-reported everyday functioning: an ecological momentary assessment approach. *Clin Neuropsychol*. 2020;34(4):678–699.
  102. Hernandez R, Hoogendoorn C, Gonzalez JS, et al. Reliability and validity of noncognitive ecological momentary assessment survey response times as an indicator of cognitive processing speed in people's natural environment: intensive longitudinal study. *JMIR mHealth and uHealth*. 2023;11:e45203.
  103. Vesel C, Rashidisabet H, Zulueta J, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a BiAffect iOS study. *J Am Med Inform Assoc*. 2020;27(7):1007–1018.
  104. Romanowska S, Best MW, Bowie CR, et al. Examining the association of life course neurocognitive ability with real-world functioning in schizophrenia-spectrum disorders. *Schizophr Res Cogn*. 2022;29:100254.
  105. Islam MA, Habtewold TD, van Es FD, et al. Long-term cognitive trajectories and heterogeneity in patients with schizophrenia and their unaffected siblings. *Acta Psychiatr Scand*. 2018;138(6):591–604.
  106. Granholm E, Ben-Zeev D, Fulford D, Swendsen J. Ecological Momentary Assessment of social functioning in schizophrenia: impact of performance appraisals and affect on social interactions. *Schizophr Res*. 2013;145(1-3):120–124.
  107. Gard DE, Sanchez AH, Cooper K, Fisher M, Garrett C, Vinogradov S. Do people with schizophrenia have difficulty anticipating pleasure, engaging in effortful behavior, or both? *J Abnorm Psychol*. 2014;123(4):771–782.
  108. Campellone TR, Sanchez AH, Fulford D, Kring AM. Defeatist performance beliefs in college students: transdiagnostic associations with symptoms and daily goal-directed behavior. *Psychiatry Res*. 2019;272:149–154.
  109. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. *Int J Methods Psychiatr Res*. 2017;26(4):e1562.
  110. Moore RC, Parrish EM, Van Patten R, et al. Initial psychometric properties of 7 NeuroUX remote ecological momentary cognitive tests among people with bipolar disorder: validation Study. *J Med Internet Res*. 2022;24(7):e36665.
  111. Parrish EM, Kamarsu S, Harvey PD, Pinkham A, Depp CA, Moore RC. Remote ecological momentary testing of learning and memory in adults with serious mental illness. *Schizophr Bull*. 2021;47(3):740–750.
  112. Depp CA, Kamarsu S, Filip TF, et al. Ecological momentary facial emotion recognition in psychotic disorders. *Psychol Med*. 2022;52(13):2531–2539.
  113. Nicosia J, Aschenbrenner AJ, Balota DA, et al. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. *J Int Neuropsychol Soc*. 2023;29(5):459–471.
  114. Moore RC, Ackerman RA, Russell MT, et al. Feasibility and validity of ecological momentary cognitive testing among older adults with mild cognitive impairment. *Front Digit Health*. 2022;4:946685.
  115. Moore RC, Paolillo EW, Sundermann EE, et al. Validation of the mobile verbal learning test: illustration of its use for age and disease-related cognitive deficits. *Int J Methods Psychiatr Res*. 2021;30(1):e1859.
  116. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. *Assessment*. 2018;25(1):14–30.
  117. Thompson LI, Harrington KD, Roque N, et al. A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimers Dement (Amst)*. 2022;14(1):e12283.
  118. Singh S, Strong R, Xu I, et al. Ecological momentary assessment of cognition in clinical and community samples: reliability and validity study. *J Med Internet Res*. 2023;25:e45028.
  119. Silva AM, Limongi R, MacKinley M, Ford SD, Alonso-Sánchez MF, Palaniyappan L. Syntactic complexity of spoken language in the diagnosis of schizophrenia: a probabilistic Bayes network model. *Schizophr Res*. 2023;259:88–96.
  120. Liebenthal E, Ennis M, Rahimi-Eichi H, Lin E, Chung Y, Baker JT. Linguistic and non-linguistic markers of disorganization in psychotic illness. *Schizophr Res*. 2023;259:111–120.
  121. Charalambous AP, Pye A, Yeung WK, et al. Tools for app- and web-based self-testing of cognitive impairment: systematic search and evaluation. *J Med Internet Res*. 2020;22(1):e14551.
  122. Thompson LI, Kunicki ZJ, Emrani S, et al. Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimers Dement (Amst)*. 2023;15(4):e12500.

123. Berron D, Glanz W, Clark L, *et al.* A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *NPJ Digit Med.* 2024;7(1):79.
124. Oravec Z, Harrington KD, Hakun JG, *et al.* Accounting for retest effects in cognitive testing with the Bayesian double exponential model via intensive measurement burst designs. *Front Aging Neurosci.* 2022;14:897343.
125. Novack MA, Young SR, Dworak EM, *et al.* Mobile toolbox (MTB) remote measures of executive function and processing speed: development and validation. *J Int Neuropsychol Soc.* 2024:1–9.
126. Check B. Available at: <https://braincheck.com/platform/>
127. vital Db.
128. Health L. Available at: <https://linushealth.com>
129. Cognition C. Available at: <https://cambridgecognition.com/digital-cognitive-assessments/>
130. Neotiv. Available at: <https://neotiv.com/en>
131. Cogstate. Available at: <https://www.cogstate.com/digital-cognitive-assessment/>
132. Paolillo EM, Bomyea J, Depp CA, Henneghan AM, Raj A, Moore RC. Characterizing performance and normative standards on a suite of English-language NeuroUX mobile cognitive tests. *In Review.*
133. Parrish EM, Lin J, Scott V, *et al.* Mobile facial affect recognition and real-time social experiences in serious mental illness. *Schizophr Res Cogn.* 2022;29:100253.
134. Parrish EM, Pinkham A, Moore RC, *et al.* An ecological momentary cognitive assessment study of over-attribution of threat and suicide risk factors in people with serious mental illness. *Schizophr Res.* 2024;266:136–144.
135. Torous J, Wisniewski H, Bird B, *et al.* Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *J Technol Behav Sci* 2019(4):73–85. 4
136. Liu G, Henson P, Keshavan M, Pekka-Onnela J, Torous J. Assessing the potential of longitudinal smartphone based cognitive assessment in schizophrenia: a naturalistic pilot study. *Schizophr Res Cogn.* 2019;17:100144.
137. Shvetz C, Gu F, Drodge J, Torous J, Guimond S. Validation of an ecological momentary assessment to measure processing speed and executive function in schizophrenia. *npj Schizophr.* 2021;7(1):64.
138. Henson P, Rodriguez-Villa E, Torous J. Investigating associations between screen time and symptomatology in individuals with serious mental illness: longitudinal Observational Study. *J Med Internet Res.* 2021;23(3):e23144.
139. Depp CA, Bashem J, Moore RC, *et al.* GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *NPJ Digit Med.* 2019;2:108.
140. Rough IM, James SH, Gonzalez CM, *et al.* Geolocation as a digital phenotyping measure of negative symptoms and functional outcome. *Schizophr Bull.* 2020;46(6):1596–1607.
141. Daniel DG, Cohen AS, Velligan D, *et al.* Remote assessment of negative symptoms of schizophrenia. *Schizophr. Bull. Open.* 2023;4(1):sgad001.
142. He-Yueya J, Buck B, Campbell A, *et al.* Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. *npj Schizophr.* 2020;6(1):35.
143. Buck B, Scherer E, Brian R, *et al.* Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. *Schizophr Res.* 2019;208:167–172.
144. Lane E, D’Arcey J, Kidd S, *et al.* Digital Phenotyping in adults with schizophrenia: a narrative review. *Curr Psychiatry Rep.* 2023;25(11):699–706.
145. Miller ML, Rough IM, Strauss GP, Harvey PD. Remote digital phenotyping in serious mental illness: focus on negative symptoms, mood symptoms, and self-awareness. *Biomark Neurop.* 2022;6(100047):100047.
146. Russell MT, Funsch KM, Springfield CR, *et al.* Validity of remote administration of the MATRICS Consensus Cognitive Battery for individuals with severe mental illness. *Schizophr Res Cogn.* 2022;27:100226.
147. Harvey PD, Zayas-Bazan M, Tibirica L, Kallestrup P, Czaja SJ. Improvements in cognitive performance with computerized training in older people with and without cognitive impairment: synergistic effects of skills-focused and cognitive-focused strategies. *Am J Geriatr Psychiatry.* 2022;30(6):717–726.
148. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. *JMIR Serious Games.* 2018;6(4):e10839.
149. United States Food & Drug Agency. Digital health technologies for remote data acquisition in clinical investigations: guidance for industry, investigators, and other stakeholders. Accessed August 11, 2023.
150. Holleman GA, Hooge ITC, Kemner C, Hessels RS. The ‘Real-World Approach’ and Its problems: a critique of the term ecological validity. *Front Psychol.* 2020;11:721.
151. Suchy Y, DesRuisseaux LA, Gereau Mora M, Brothers SL, Niermeyer MA. Conceptualization of the term “ecological validity” in neuropsychological research on executive function assessment: a systematic review and call to action. *J Int Neuropsychol Soc.* 2024;30(5):499–522.
152. Suchy Y, Mora MG, DesRuisseaux LA, Niermeyer MA, Brothers SL. Pitfalls in research on ecological validity of novel executive function tests: a systematic review and a call to action. *Psychol Assess.* 2024;36(4):243–261.
153. United States Food & Drug Agency. Clinical outcome assessments (COA) qualification program. Accessed August 11, 2023.
154. European Medicines Agency. Qualification of novel methodologies for drug development: guidance to applicant. Available at: [https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants\\_en.pdf](https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf)
155. Liu Q, Zhu H, Liu C, *et al.* Application of machine learning in drug development and regulation: current status and future potential. *Clin Pharmacol Ther.* 2020;107(4):726–729.
156. European Medicines Agency. Reflection paper on the use of artificial intelligence in the lifecycle of medicines. Accessed August 11, 2023.