

Preliminary Psychometric Comparisons of Remote-Televideo and Face-To-Face Administration of a Commonly Applied Neurocognitive Assessment Battery

Aaron S. Kemp^{1,2}; Kirstan N. Gooch²; Joisabel L. Goldberg²; Sitha Bun²; Dimitrios Gripeos^{1,3}; Christopher Reist^{1,3}; Barton W. Palmer^{4,5}; Mark Bondi^{4,5}; James P. O'Halloran^{2*}

¹Psychiatry and Human Behavior, University of California, Irvine School of Medicine; ²NeuroComp Systems, Inc., Irvine, CA;

³Veterans Administration Healthcare System, Long Beach, CA; ⁴Psychiatry, University of California, San Diego; ⁵Veterans

Medical Research Foundation, San Diego, CA; *Corresponding author: jpo@neurocomp.com; 949-752-7560



2009

This poster is presented in columns for online reading.
You may also see the poster in its [original format](#).

Abstract

Background: Neurocognitive assessment has become an integral component of controlled clinical trials of candidate “cognitive-enhancing” treatments for a wide range of CNS disorders. However, manual administration of large, paper-based, neurocognitive assessment batteries is often inefficient, error-prone, and inconsistent across multiple sites. Existing computerized testing systems are also limited in both the assessment instruments available and the range of impairments that can be accommodated with the subject sitting alone at a single display. Therefore, a unique dual-display computerized testing system was developed, with funding from the National Institute of Mental Health (NIMH), that integrates (rather than replaces) the examiner for computerized administration of standard neurocognitive assessment batteries (O’Halloran et al., 2007 & 2008). This novel, dual-display, configuration has now been extended to support two-way, wireless televideo communications for remote administration by an expert examiner at a distal location.

Purpose: The purpose of the current study was to evaluate the psychometric feasibility of this novel configuration by comparing the concurrent validity and test-retest reliability of computerized, remote-televideo (RT) administration of a representative battery of common neurocognitive assessment instruments with traditional, face-to-face (FF) administration of the same battery on paper.

Methods: The neurocognitive battery was administered to 25 healthy subjects with no history of psychiatric diagnoses, 5 patients with schizophrenia, and 10 patients with Mild Cognitive Impairment (MCI) via both methods ~14 days apart with the order of administration counterbalanced across participants.

Results: Intraclass Correlation Coefficient (ICC) comparisons of concurrent validity between RT and FF batteries yielded highly significant measures of agreement for all tests and only one significant mean difference was found between the methods using paired-samples t-test comparisons. The ICCs for test-retest reliability were also highly significant for all tests compared.

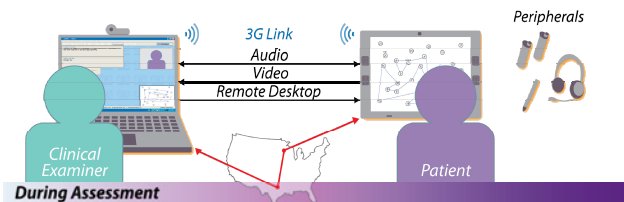
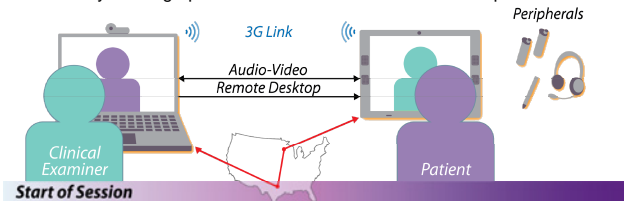
Conclusions: While further validation within the specific patient populations for which application is intended is still on-going, the current, preliminary results support the psychometric feasibility of administering a computerized battery of commonly applied neurocognitive assessment instruments via remote, televideo interactions with an expert examiner at a distal location.

Methods

Participants included patients diagnosed with mild cognitive impairment (N=10) or schizophrenia (N=5), and healthy subjects with no history of psychiatric diagnoses (N=25). A brief battery of neurocognitive assessment instruments was administered via both RT (computerized) and FF (paper-based) methods by a trained psychometrist on two separate visits ~14 days apart with the order counterbalanced across participants. For half of the RT sessions the participants were located in one facility and tested from another; for the other half the participants were located in their own residence. Concurrent validity comparisons were conducted using Intraclass Correlation Coefficients (ICC) to assess absolute agreement and Pearson’s *r* correlation coefficients to assess covariance. Test-retest reliability comparisons were also conducted using ICCs to assess the relative consistency of the data collected between the first and second visits, separately for each administration mode.

Remote-Televideo Configuration

A unique, dual-display, testing system (**CMINDS**[®]) was utilized at the patient's location and a separate laptop was used by the remote examiner to establish and maintain the two-way, wireless televideo communications between the patient and examiner throughout the assessment session and enable control of the test system's graphical user interface via remote desktop connection.



High-speed, very low latency, wireless 3G (and eventually, 4G) connectivity permits the subject and examiner to freely interact throughout a session via secure, high-resolution, audio-video interface, as is critical for tests requiring interactive scoring of verbal report or behavioral performance. Accordingly, the examiner is able to control the pacing of the test session, repeat and/or supplement computerized instructions (to ensure understanding of the task demands), immediately score verbal report or overt behavior in real time and pause or terminate a test, as may be required due to unforeseen interruptions or inability of a patient to perform a given task. Thus, the system is well suited for application in impaired populations that may require ongoing monitoring and supplemental instruction. The **CMINDS**[®] test set consists of computerized versions of established instruments, commonly applied in clinical trials.

Results: Concurrent Validity

As can be seen in **Table 1**, concurrent validity comparisons yielded highly significant ($p < .001$) measures of absolute agreement on the ICCs for all tests compared. Pearson's r coefficients also yielded highly significant measures of covariance for all comparisons. The paired samples t-tests indicated only one significant mean difference between FF and RT. For Symbol-Digit, it was discovered that the lag between the subject's verbal report, the examiner's scoring of such, and the subsequent advancement of the screen via remote desktop contributed to lower scores. Accordingly, during Phase II of this SBIR project, the system will be reconfigured to permit real-time synchronization between distributed instantiations of **CMINDS**[®] and a dedicated hardware platform will be developed to ensure uniformity in system connectivity. This will negate any reliance on off-the-shelf remote-desktop utilities and will confer greater control than the inherent variability of commercial hardware platforms.

Table 1: Concurrent Validity Comparisons (n=40)

	Means (sd)		Correlation Coefficients	
	FF	RT	ICC (sig.)	Pearson's r (sig.)
NAART Total Correct	40.6 (9.6)	40.2 (8.9)	0.94 (<.001)	0.89 (<.001)
AVLT Immediate Recall	24.1 (7.0)	23.8 (7.0)	0.91 (<.001)	0.84 (<.001)
AVLT Delayed Recall	8.0 (4.3)	7.8 (4.0)	0.87 (<.001)	0.78 (<.001)
AVLT Delayed Recognition	12.2 (2.8)	11.9 (2.8)	0.84 (<.001)	0.73 (<.001)
Symbol-Digit Substitution	50.4 (16.1)	42.8 (15.2)	0.86 (<.001)	0.76 (<.001)
Clock Drawing	4.83 (0.6)	4.74 (0.7)	0.78 (<.001)	0.69 (<.001)
Trails A Total Time	37.8 (22.6)	40.4 (24.7)	0.79 (<.001)	0.71 (<.001)
Trails B Total Time	78.8 (41.6)	82.7 (44.7)	0.84 (<.001)	0.73 (<.001)
Letter-Number Span	10.8 (3.2)	10.3 (3.4)	0.91 (<.001)	0.84 (<.001)
Letter Fluency	45.6 (11.7)	46.5 (12.3)	0.86 (<.001)	0.76 (<.001)
Category Fluency	48.5 (12.7)	50.7 (13.4)	0.91 (<.001)	0.83 (<.001)

Results: Test-Retest Reliability

As can be seen in **Table 2**, test-retest reliability comparisons yielded highly significant ($p<.001$) measures of consistency on the ICCs for all tests compared. Thus, the data collected via RT administration exhibits substantially equivalent reliability as that collected via traditional, FF administrations.

Table 2: Test-Retest Reliability Comparisons (n=40)

	Face-to-Face			Remote-Video		
	Visit 1	Visit 2	ICC	Visit 1	Visit 2	ICC
AVLT Im. Recall	24.1 (7.0)	24.8 (7.2)	0.91	23.8 (7.0)	24.1 (7.4)	0.88
AVLT Delay Recall	8.0 (4.3)	8.1 (4.8)	0.87	7.8 (4.0)	7.9 (4.2)	0.91
AVLT Delay Recog.	12.2 (2.8)	12.5 (2.9)	0.84	11.9 (2.8)	12.0 (2.7)	0.82
Symbol-Digit Subst.	50.4 (16.1)	52.8 (17.0)	0.96	42.8 (15.2)	43.3 (15.7)	0.89
Clock Drawing	4.83 (0.6)	4.76 (0.6)	0.82	4.74 (0.7)	4.79 (0.8)	0.86
Trails A Total Time	37.8 (22.6)	34.3 (18.8)	0.90	40.4 (24.7)	37.4 (26.7)	0.88
Trails B Total Time	78.8 (41.6)	72.7 (47.7)	0.91	82.7 (44.7)	75.8 (42.3)	0.90
Letter-Number Span	10.8 (3.2)	11.2 (2.9)	0.93	10.3 (3.4)	10.9 (3.3)	0.91
Letter Fluency	45.6 (11.7)	46.8 (12.3)	0.88	46.5 (12.3)	48.2 (13.3)	0.88
Category Fluency	48.5 (12.7)	48.2 (12.9)	0.91	50.7 (13.4)	52.4 (14.1)	0.89

Conclusions

The current results support the psychometric feasibility of administering a computerized battery of common neurocognitive assessment instruments via remote, televideo interactions with an expert examiner at a distal location. While still requiring further validation within various patient populations and customization of a dedicated hardware platform, the current, preliminary findings suggest potential for cost-effective utilization in clinical trials, on which improved reliability in outcome measurement directly translates into reduced sample size requirements, associated cost savings, and a minimization of the risks of placebo exposure for participants in randomized, controlled trials.