STANDARDISATION AND CROSS-VALIDATION STUDY OF COGTEST – AN AUTOMATED COGNITIVE BATTERY FOR USE IN CLINICAL TRIALS

Tonmoy Sharma [1], Robert Bilder [2]

1 Clinical Neuroscience Research Centre, Dartford, Kent, United Kingdom &
2 Department of Psychiatry, University of California at Los Angeles

ABSTRACT

We sought to standardise and cross-validate Cogtest, an automated neurocognitive battery with electronic data capturing ability. 200 individuals took part in 2 studies, one aimed at assessing concurrent validity with paper-pencil (P&P) measures, one establishing norms for individual Cogtest endpoints that comprise a Neurocognitive Global Score (NGS).

The first study (N=75) involved parallel administration of Cogtest, and a battery of P&P tests, in 2 sessions. Both sets of instruments were presented in counterbalanced order at both sessions, with testing at 0 and 4 weeks (+/- 3 days) control. Four forms of the Cogtest battery were counterbalanced. The 75 subjects were stratified from age 20 to 60. The second study (N=120) involved 2 "baseline" sessions and one follow-up session, and 10 male, 10 female participants in each of 6 age bands from age 13 to 69.

Study 1 revealed sex differences, with men more accurate on spatial processing but women more accurate in face recognition memory. Correlating Cogtest measures with P&P tests revealed correlations ranged r=.3 to r=.7. Study 2 revealed a curvilinear age effect on NGS and individual domains; lower scores in the 13-19 year age group, highest scores in the 20-29 year group, and monotonically decreasing scores through the 6th decade. Both studies revealed that test-retest stability up to 4 weeks was in keeping with published results for P&P, with test-retest reliability coefficients in the range of ~0.4 to ~0.9.

This study shows Cogtest has similar psychometric properties to P&P tests, and offers the advantages of computerisation and audit trail.
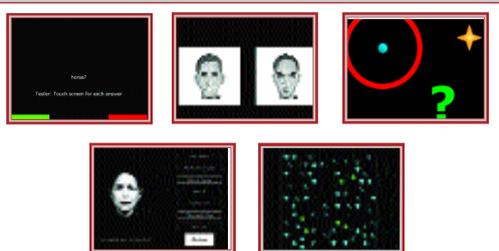
INTRODUCTION

Cognitive function enables processing of information and organises our thoughts and actions. Cognitive performance can be described and measured in different domains including memory, attention, motor function, general executive functioning, and perpetual organization. Deficits in cognition are part of the etiology of several neuropsychological disorders, e.g., schizophrenia, Alzheimer's disease, etc.

OBJECTIVES

Study 1: To standardise the Cogtest battery (Cogtest plc, London; www.cogtest.com) against P&P equivalent tests.

Study II: To establish means and standard deviations for the Cogtest endpoints on cognitively healthy individuals that will feed into the Neurocognition Composite Score (NCS) to enable:

a) Computation of a NCS from subject screening data which will serve as a baseline reference in future studies.

b) Characterisation of severity of cognitive deficit at baseline of the study sample;

c) To serve as a reference for inter-study comparisons of NCS.;

d) Interpretation of NCS=0.0 as "normal" on all tests.

METHODS

All participants were administered the following tests, to test-retest the reliability of the Cogtest battery:

- Cogtest Workstation Orientation
- Continuous Performance Test – AX Version
- Continuous Performance Test – Identical Pairs Version
- Continuous Performance Test Flanker Version
- Strategic Target Detection Test – 4-shape Version
- Object Working Memory Test
- Spatial Working Memory Test
- Facial Memory Test (immediate and delayed)
- Word List Memory Test (immediate and delayed)
- Symbol Digit Substitution
- Competing Programs Test
- Auditory Digit Span
- Set Shifting Test
- Simple Go-No Go test
- Auditory Letter-Number Sequencing
- Emotion recognition test (PEAT)

The average time to complete the battery was approximately 1 hr.

Figure 1: Screenshot Examples of Cogtest battery tests

Paper and Pencil Tests for the Cogtest Standardization Study

Test	Description	Time (minutes)
AVLT	Verbal list learning test	15 + 5 = 20
Digit Span (WAIS III)	Digit recall and working memory	5
Letter-Number Span (WAIS III)	Working memory	5
Coding (WAIS III)	Attention and processing speed	4
WCST	Executive functions	20
TMT	Executive functions	6
Verbal Fluency		
(COWAT, Animal Naming)	Executive functions	6
Finger Tapping Test	Motor speed	5
Grooved Pegboard Test	Motor dexterity	5
Vocabulary (WAIS III)	Word knowledge	10
MMSE – Mini Mental State Exam	General ability	12
ADAS-Cog	General ability	20 (two highest age strata only)

Study I: The overall design involved parallel, counterbalanced administration of both the Cogtest test battery and a battery of P&P tests. There were two test sessions, with both sets of instruments presented at both sessions, with time between testing of 4 to 8 weeks (paralleling the duration of typical clinical trials). There were 4 alternate forms of each test in the Cogtest battery. Subjects were assigned one of these at each session.

Study II: This methodology study followed a single centre, parallel group design, and was conducted in male and female adult subjects. One hundred and twenty (6 groups of 20) subjects were required. Groups were defined as follows:

Endpoints and Criteria for Evaluation:

Collection of results for the cognition tests listed in Methods:

Group	Age Range (yrs)	Males	Females
1	13-19	10	10
2	20-29	10	10
3	30-39	10	10
4	40-49	10	10
5	50-59	10	10
6	60-69	10	10

RESULTS

Correlations of computerised test results with age were nonsignificant, except for WO reaction time, which was positively correlated with age (r=.27, p=.027, n = 68). It should be recognised that this was despite a trend for the older subjects to have higher scores on measures usually associated with higher levels of premorbid ability (Vocabulary: r=.47, p=.001, n=71; phonemic fluency: r=.44; p<.001; n=70; WRAT Reading: r=.29, p=.015, n=71; Mini-Mental State Exam: r=.26, p=.030, n=70). Partial correlations of age with computerised test scores, controlling for vocabulary, revealed that older subjects were more accurate in target selection (WO distance: r=-.28, p=.025, df=68), and also in the Flanker CPT (r=.28, p=.023, df=66).

Significant sex effects were found on both P&P and computerised tests. Specifically, men were more accurate on the Flanker CPT, were faster on the Flanker CPT, Set Shifting Test and Tapping Speed Test, and more accurate on the Spatial Working Memory Test. In contrast, women were more accurate on the Face Memory Test, and faster on the computerised Word List Memory test.

SUMMARY OF RESULTS

Study II:

There were no significant gender effects except for the total errors endpoint in the competing programs test. In general, this test was found to be problematic as there were a number of individuals that failed to complete the practice and never progressed to the actual test. The declarative memory score does not follow the pattern of learning as do the other domains. **Figure 2.** shows the composite scores across age groups for the test session responses standardised to the overall means.

Figure 2. Age Effect on Composite Scores
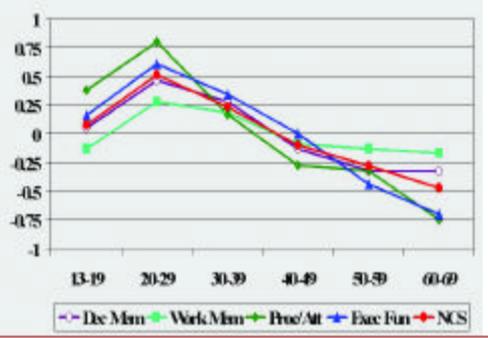
DISCUSSION

Study I : The test-retest stability data shows the range of correlations (r) for the test retest which ranged from 0.38 (Face matching) to 0.84 (word list memory), with an Overall Test Battery Mean (OTBM) test retest of 0.57. There were exceptions of the object- and spatial delayed- working memory of showing 0.29. This suggests that this combination of tests has good reliability, both for individual tests and for the battery overall. Our results suggest that the battery has adequate consistency for test-retest comparison at about 4 weeks.

Study II gives rise to two recommendations 1) Using the age-specific means for standardisation in creating composite scores and 2) incorporating at least two practice sessions before establishing the study baseline for these tests.

This study validates the use of the Cogtest computerized battery in studies of MCI and AD

REFERENCES

Button, E. (1996). Validation and invalidation. In J.W. Scheer & A. Catina (Eds.) Empirical Constructivism in Europe: The Personal Construct Approach. Giessen: Psychosozial-Verlag, pp. 142-148.

Habel, U., Gur, R.C., Mandal, M.K., Salloum, J.B., Gur, R.E., & Schneider, F. (2000) Emotional processing in schizophrenia across cultures:standardised measures of discrimination and experience. Schizophrenia Research 42:1, 57-66

Kohler, C.G., Bilker, W., Hagendoorn, M., Gur, R.E., & Gur, R.C. (2000) Emotion recognition deficit in schizophrenia: Association with symptomatology and cognition. Biological Psychiatry 48: 127-136.

Poole, J.H., Tobias, F.C., & Vinogradav, S. (2000) The functional relevance of affect recognition errors in schizophrenia. Journal of the International Neuropsychological Society 6:6, 649-658

Bilder, R.M., Turkel, E., Lipschutz-Broch, L., Lieberman, J.A. (1992). Antipsychotic medication effects on neuropsychological functions. Psychopharmacology Bulletin, 28, 353-366

Lipschutz-Broch, L., Bilder, R.M., Turkel, E., Reiter, G., Lieberman, J.A. (1993, April). A chronometric analysis of executive deficits in acute and stable phases of first episode schizophrenia. IVth International Congress on Schizophrenia Research, Colorado Springs, CO

Lipschutz-Broch, L., Bilder, R.M., Turkel, E., Reiter, G., Lieberman, J.A. (1993, May). Validation of new measures of executive dyscontrol. Fourth-Eighth Annual Convention of the Society for Biological Psychiatry, San Francisco, CA

Table1. Descriptive Statistics for Computerised Tests at Baseline

Table2. Descriptive Statistics for Paper and Pencil Tests at Baseline