2001

The Remote Neuropsychological Assessment -Category Test: Development and Validation of a Computerized, Internet -Based Neuropsychological Assessment Measure.

Jeffrey Nicholas Browndyke
Louisiana State University and Agricultural & Mechanical College

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

THE REMOTE NEUROPSYCHOLOGICAL ASSESSMENT - CATEGORY TEST: DEVELOPMENT AND VALIDATION OF A COMPUTERIZED, INTERNET-BASED NEUROPSYCHOLOGICAL ASSESSMENT MEASURE

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Jeffrey Nicholas Browndyke
B.S., University of Memphis, 1992
M.A., Louisiana State University, 1997
May 2001

"Man is still the most extraordinary computer of all."

John F. Kennedy
May 21, 1963

ACKNOWLEDGEMENTS

I would like to extend my sincerest gratitude towards the members of my dissertation committee for their feedback and guidance during the inception, planning, and completion of this project. The direction of my graduate mentor and dissertation chair, Wm. Drew Gouvier, Ph.D., deserves special mention. His patience, understanding, and encouragement in this research pursuit have been unflagging, and I owe him a debt of gratitude for allowing me to formulate and foster a personal research focus during my graduate school training. The mentorship and critical research aid provided during my graduate program training by William F. Waters, Ph.D. is most appreciated. My understanding of psychophysiology and neuropsychology has been enriched by your tutelage. Similarly, the critical thinking skills and grounding in neuroscience, facilitated by Mike Hawkins, Ph.D., have been invaluable to my training. Recognition of important clinical issues, the establishment of methodological rigor, and statistical knowledge have been shared generously by Donald A. Williamson, Ph.D. Not only have I grown as a scientist-practitioner under each of you, but I have also gained valuable personal examples of scientific integrity, rigor, and insight, which I hope will guide my career in the years ahead. Additionally, I would like to extend a warm thank you to Sarah Pierce, Ph.D., who graciously gave of her time and effort to serve on my dissertation committee.

I would be greatly remiss in not mentioning the primary influence in my life – my family. Their unwavering encouragement, support, and faith bolstered my reserve and steadied my focus during this project. I sincerely thank my parents for their sacrifices and support over the years, and to my brother and sister, I toast our friendship and collective devotion. Additionally, I would like to recognize Karen Tucker for being there during the tough times and renewing my faith in life, love, and happiness. I look forward to traveling life's road with you and profiting from our shared experience. I am truly blessed to have each of you in my life and would be spiritually impoverished without your presence.

Lastly, I would like to thank the tireless and charitable efforts of those who have aided in this project. William Malone for his instrumental work and consultation in computer programming and multimedia design, without which the RNA-CT would still be just an idea on the drawing board. The data collection efforts of John Denning and Ashlie Albert insured the timely completion of this project and underscored their dependability and insight. And, finally, to all my friends in and out of graduate school for their encouragement, commiseration, and unyielding criticism, I raise my glass to your lives and wish you all the best.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In the current investigation, a prototype for the remote assessment of cognitive skills, the "remote neuropsychological assessment (RNA) model," was proposed, and the development and validation of a computerized, Internet-based neuropsychological assessment measure was undertaken to demonstrate the utility and effectiveness of this untapped model of assessment delivery. The Remote Neuropsychological Assessment - Category Test (RNA-CT), a test of abstract concept formation administered via the World Wide Web (WWW), was developed and contrasted with a conventional, non-computerized version of the measure – the Booklet Category Test (BCT). Traditional and novel measurement variables from both measures were compared in a randomized group design of normal college-educated subjects in an attempt to demonstrate equivalence between the conventional assessment and RNA model. Comparison of the equivalence between administration types suggests significant convergence for total error, subtest error variables, and internal factor structure between measures. However, differences in the amount of variance allotted to the factor structures of the subtest error variables suggests that, while equivalent in their outcome variables and internal test structures, the measures tend to differentially weigh their component constructs (i.e., congeneric equivalence). In contrast to the equivalence noted between measures on their primary variables, significant differences between administration types were noted for the majority of the response timing

variables (e.g., total test time and total time per subtest). The results of the timing variable comparisons suggest that the RNA-CT may prove to be a more time efficient test. However, a significant positive relationship between timing and error variables was noted for the RNA-CT; a relationship that was not fully supported in the BCT analyses. Future research is needed before the significance or utility of RNA-CT and BCT timing variables can be established.

Overall, the results from the current investigation point toward a high degree of convergent and construct validity between RNA and conventional administrations of a neuropsychological assessment measure in a normal, college-educated population. The theoretical and practical implications of employing a RNA model are discussed and suggestions are provided for the future evaluation and implementation of RNA in modern clinical practice.

INTRODUCTION

Based upon the findings of a recent random telephone survey of 3,493 adults, the Pew Research Center reported that last year in the United States, the Internet was used by 56% of the adult population, accounting for approximately 104 million regular Internet users (Pew Internet & American Life Project, 2001). This number of Internet users is up considerably from the 20% reported by the Pew Research Center in December of 1996 (Mund, 1996) and serves to underscore the ascendancy of the Internet in modern culture. And, while the majority of the Internet is being increasingly used for email communication (e.g., 93% of current users sent or received email at least once every few weeks; Pew Internet & American Life Project, 2001) and WWW browsing (e.g., 51% of users accessed WWW pages the day prior to their survey participation; Mund, 1996), there remains a vast untapped resource for the administration and delivery of neuropsychological and psychological services via the Internet and other telecommunication technologies. Only recently has there been a systematic effort by some researchers to explore this virtually untapped service domain, but all of these recent efforts have been limited to the remote application of "telepsychiatry" (Allen & Wheeler, 1998; Baer et al., 1995; Elford et al., 2000) and behavioral health initiatives (Lange, van de Ven, Schrieken, Bredeweg, & Emmelkamp, 2000; Tate, Wing, & Winett, 2001; Zarate et al., 1997). Of the handful of research efforts that have indicated the use of at least one cognitive assessment measure in a remote

fashion (Ball & McLaren, 1997; Ball & Puffett, 1998; Ball, Scott, McLaren, & Watson, 1993; Ball, Tyrrell, & Long, 1999; Harvey, Roques, Fox, & Rossor, 1998; Lee et al., 2000; Montani et al., 1997; Monteiro et al., 1998; Nesselroade, Pedersen, McClearn, Plomin, & Bergeman, 1988) or the use of remote interventions for brain injury or epilepsy rehabilitation (Hufford, Glueckauf, & Webb, 1999; Schopp, 2000; Schopp, Johnstone, & Merveille, 2000), all have been limited to the use of telephony or televideo technologies.

To date, there has been no systematic study reported that has investigated the feasibility or effectiveness of Internet technology in the administration of a neuropsychological assessment measure. The following investigation was viewed as an initial step towards the integration of neuropsychological measurement techniques and remote cognitive assessment via Internet-based technology. The Category Test (CT; Halstead, 1940, 1947; Halstead & Settlage, 1943; Halstead & White, 1950; Reitan, 1969; Reitan & Wolfson, 1985), a well-known and widely used neuropsychological assessment paradigm tapping abstract conceptualization skills and particularly sensitive to brain dysfunction, was selected as a model for the construction of, what is believed to be, the first Internet-based neuropsychological assessment measure and the first attempt to empirically validate such an instrument. A model for the effective use of Internet technology in the remote administration of neuropsychological assessment was proposed [i.e., the remote neuropsychological assessment (RNA) model], which guided the construction

and planning of the measure studied in this project, the Remote Neuropsychological Assessment – Category Test (RNA-CT; see Appendix B).

What follows is a brief exploration of computer technology in cognitive assessment, then a discussion of the use of telecommunications and Internet technology by the field of neuropsychology, and completed by a general review of the literature related to the CT[1].

[1]An abbreviation using the letters, CT, in isolation, denotes a general reference to the Category Test assessment paradigm with all of its numerous iterations; whereas, other abbreviations used in this manuscript refer to the specific versions of the CT (e.g., HCT, HRCT, BCT, ICT, RNA-CT).

REVIEW OF LITERATURE

Computer Technology in Cognitive Assessment

In an era of shrinking mental health funds, clinicians often find themselves beset by the need for cost efficient treatment and assessment implementations. One of the pathways this cost efficiency movement travels is the automation and standardization of mental health contact via computerization. Many mental health fields are beginning to adopt the use of computers as a method to maintain professional control while simultaneously placating the cost efficiency demands generated by the managed care operations (MCOs; Jerome et al., 2000). Unfortunately, the field of clinical neuropsychology has been relatively slow to make the move towards computerization, and in a field where clinical contact time strains the limits of what is reimbursable, a failure to follow this trend may ultimately be injurious to the profession. Understandable and practical limitations of computerized assessment have been voiced as factors responsible for the reticence to adopt computerization within clinical neuropsychology on a large scale (Kane & Kay, 1992; Schatz & Browndyke, 1999). However, as will be discussed, these limitations and concerns are self-imposed and not current with what today's computer technology will allow.

In the age of information, the increased availability and reduced cost of the computer has opened up a potential new era in neuropsychological assessment. The clinical use of computerized assessment, though spanning back more than two decades (Beaumont, 1975), is still in its early stages, and

current computer hardware and software has not been able to fully address all the needs of clinicians and patients. Nevertheless, the potential contribution of the computer to neuropsychological assessment is large and the continued development of these automated procedures in inevitable, particularly in an era where managed-care expediency and time-cost analyses fuel many aspects of clinical decision making.

Computerized testing offers many advantages over conventional neuropsychological testing with respect to test administration, response monitoring, and scoring. The computer is able to provide precise control over the presentation of test stimuli, thereby potentially increasing test reliability. In a computerized test, software controls the visual and auditory stimulus characteristics. Programs can adaptively control the order, number, presentation rate, and/or complexity of items. The computer is also capable of controlling contrast intensity. There is also the option of presenting degraded auditory and visual stimuli. Many of these stimulus control advantages simply cannot be achieved by conventional testing (Kane & Reeves, 1997; Mead & Drasgow, 1993).

Computers are also especially well suited for repeated testing in both clinical and research settings (Englund, Reeves, Shingledecker, Wilson, & Hegge, 1987). Neuropsychologists frequently encounter patients who are referred for repeat testing. The follow-up evaluation is generally requested to help the treating physician gauge the patient's response to medication or

medical treatment or progression of a disease process. Unfortunately, many conventional neuropsychological tests are not designed for re-administration. These tests tend to show substantial practice effects; confounds that markedly diminish their sensitivity to changes in brain functioning and render them less effective tools for monitoring changes in patient performance. In contrast, the computer is capable of generating multiple forms of a test. This is important not only in producing parallel forms, but also in generating a stable baseline against which to evaluate change. The feature of assessment repeatability has led to the inclusion of computerized tests in specialized clinical and research settings that require that test batteries be repeated following relatively short time intervals.

In addition to being adept at test administration, the computer keeps a superb accounting of test responses. The computer is capable of scoring and recording the accuracy and speed of each response. Upon the completion of a test, the program can calculate the relevant statistics for the test and then report the results in a variety of different formats. The computer's accuracy in recording and scoring the examinee's responses cannot be consistently matched by the human administrator (Kane & Reeves, 1997).

The properly programmed computer is not only free from stimulus presentation error, scoring error, and experimenter bias, it continues to administer tests exactly according to standardization procedures without looking for abbreviated or short forms. These improvements in standardization

allow for easier comparison of data obtained from different sites and by different examiners.

In spite of the advantages discussed above, computerized testing has limitations and drawbacks. Among the most serious deficiencies of some existing test software has been the use of inaccurate timing procedures (D. Chute, personal communication, November 1997), the use of poorly designed human-computer interfaces, the lack of usable reports and data sets, and the failure to meet established testing standards (American Psychological Association, 1986). Even for well-designed and innovative programs with accurate timing and standardization, at present the computer has only limited capability for assessing expressive language skills, and as a result, falls short in the sophistication needed to allow for testing of visual confrontation naming, expressive speech, oral reading, or repetition. In short, the computer is not yet capable, with a high degree of accuracy, of assessing all aspects of neurocognitive functioning that may be required for comprehensive neuropsychological assessment (Webster & Compeau, 1996), but this inability is predicted to be short lived as computers improve in function and speed.

Another drawback of computerized testing is that it may result in a reduction in the amount of interaction taking place between the examiner and the examinee (Space, 1981). A skilled examiner is often capable of coaxing the examinee to complete testing and to stay motivated. The skilled examiner can detect when the subject is ignoring or only partially reading test instructions.

Unlike the human test administrator, the computer is deficient in the aforementioned assessment factors and lacks the ability to provide sincere expression of compliments, criticisms, and encouraging comments (W. D. Gouvier, personal communication, 1996).

Additional difficulties in computerized assessment are more related to human bias and misunderstanding of the computer, rather than to the apparatus itself. Poorly designed applications inflate the perception of difficulties with computerized assessment. Subjects may fail to read instructions if they are poorly written. Some programs use the computer as a high-tech workbook or automated slide projector. These errors and non-innovative uses of computer technology demonstrate a lack of creativity and understanding of computers, and should not reflect upon the computer itself. Lastly, although clinicians have become familiar with word-processing and financial software packages, there are many clinicians who have at least a mild degree of computer phobia, which taints their perception of the usefulness of computerized assessment (Rosen, Sears, & Weil, 1992). These personal fears appear to have fueled the misperception that patients are largely reticent to interact with computers; a perception that does not appear to be substantiated by the literature (Burke, & Normand, 1987; Hile & Vieweg, 1996; Kane & Kay, 1992; Kane & Reeves, 1997).

Benefits and drawbacks not withstanding, the full potential of computerized assessment has yet to be realized. Existing computerized

assessment programs are largely outdated and do not incorporate the hardware and software advances of recent years. The ubiquitous nature of CD-ROM technology and the data storage capacity afforded by current computer technology, allow for true multimedia presentation of test stimuli and permit simulation of a wide range of environments. Current computer audio technology allows for the presentation of sophisticated verbal and other auditory test items, which may circumvent problems caused by patients' failing to read task instructions. And, while these advanced multimedia technologies exist, they have yet to be fully realized in computerized neuropsychological assessment, but they are starting to show prominence in related mental health fields, such as psychiatry and health psychology (Huang & Alessi, 1996; Workman, 1996).

The Integration of Telecommunications Technology and Neuropsychology

A logical extension of the move towards computer technology in cognitive assessment and treatment is the incorporation of telecommunication modalities (e.g., telephony, televideo and video-conferencing, and Internet technology, such as email and the WWW) to enhance the reach and scope of the clinician and researcher (Williams & Browndyke, 1996, November). Of the current telecommunication technologies, telephony and televideo have been the most studied, though by no means in an exhaustive fashion.

In a review of telephony applications, Ball and McLaren (1997) highlight the applicability of remote screening measures for dementia diagnosis and

related neurological conditions (Desmond, Tatemich, & Hanzawa, 1994; Harvey et al., 1998; Kent & Plomin, 1987). Ball and McLaren note that the use of telephony for cognitive evaluation and screening has the inherent advantage of the commonality and ease of use. Additionally, interventions that employ telephony evaluation have proved to be cost-effective means for the gross determination of cognitive state and abilities. However, Ball and McLaren make the valuable point that most accepted mental status screening procedures include the assessment of praxis and visuospatial abilities, both of which are not amenable to telephony measurement. This weakness was also raised by Desmond et al. (1994) and means that with telephony a less than comprehensive assessment of cognitive abilities and mental status is possible. Where telephony has succeeded, however, has been in the application of the technology for caregiver intervention and training (Smyth & Harris, 1993). Combining the use of telephony and email contact over the course of two-years in sample of caregivers of patients with dementia, Harvey et al. (1998) were able to demonstrate the effectiveness and outreach of a community-based intervention without the benefit of face-to-face contact with mental health professionals.

The inclusion of bi-directional video capabilities (aka. televideo) in cognitive assessment has been investigated as an alternative to the limited application of telephony assessment. In a study examining the efficacy and usefulness of videoconference technology in psychometric assessment scoring,

British researchers Ball et al. (1999) compared in-person, faxed-copy, and videoconferencing methods for Mini-Mental State Examination (MMSE) scoring. MMSE protocols from an elderly patient sample were recovered from the records of a community-based mental health team and scored among the various study conditions. Their study employed a PC-based videoconferencing system (PCS200, PictureTel) connected to an ISDN line at a 128 kbit/s data transfer rate. Standard commerical fax machines were employed for the faxed-copy administration mode. Three independent raters scored equal amounts of MMSE protocols over the three scoring modes. Interrater reliability estimates (kappa statistic) were collected comparing in-person scoring to faxed-copy and videoconferencing. For those items on the MMSE not amenable to televideo or faxed-copy scoring (e.g., reduction in reproduction clarity), Ball et al. applied a McNemar test with a lack of significant differences noted between incomplete data groups. Reliability estimates between the administration modes suggested that televideo and faxed-copy were reasonably approximate to in-person scoring. Televideo demonstrated relatively poor reliability for pentagram scoring ($\underline{k}$ = .47); whereas, faxed-copy pentagram scoring was reasonably good ($\underline{k}$ = .71). Ball et al. urged caution in the uncritical acceptance of psychometric material scored via televideo, citing their relatively poor reliability estimates for graphical material on the MMSE. They indicate that where scoring criteria are stringent and absent of collateral data the rate of scoring errors is likely to increase. To assist in reducing administration mode scoring

bias, Ball et al. suggest the use of materials that generate a high level of contrast (i.e., white paper and black ink) to facilitate figure feature detection.

Ball et al. (1993) found somewhat similar results with the use of a monochrome videoconferencing system in that written material from the MMSE (e.g., sentence construction and pentagram reproduction) was somewhat difficult to score. In this early study of televideo and psychometrics, their contention was that even though MMSE written material was more difficult to score, the results of a direct comparison with in-person scoring was not significant and the rescoring of televideo administered MMSE by hand did not alter impairment estimations. In a French study, the administrations of the MMSE and Clock Drawing Tests via televideo were investigated by Montani et al. (1996). Results from the Montani et al. project indicated a small, but significant difference between individuals who were administered the MMSE in-person and those receiving the MMSE remotely. Differences in the scoring of the MMSE, either in person or via televideo, were not reported.

Each of the studies into the application of televideo for cognitive evaluation highlighted the promise of this technological application, but most demonstrated consistent instrumentation difficulties (e.g., poor audio levels, video contrast difficulties impairing proper stimuli perception), which exist as the primary obstacle for the effective implementation of televideo technology. In addition, studies using televideo intervention cite a general dissatisfaction of the televideo users (both clinicians and patients) with assessment process,

despite its relative effectiveness to face-to-face contact (Elford et al., 2000; Lee et al., 2000; Schopp et al., 2000).

The application of Internet-based technology, specifically the Wide World Web, to allow for the remote assessment of neuropsychological performance was the primary interest of the current investigation. To capture the aspects of the use of telecommunication technologies in neuropsychological assessment, the term, remote neuropsychological assessment (RNA), was proposed to denote the general use of telecommunication and Internet technologies in neuropsychological assessment and practice. Psychiatry and other fields have adopted the "tele" prefix, such as "telepsychiatry" or "telehealth," but it is suggested that these terms are self-limiting, as new modes of connectivity are being invented, some of which do not necessarily involve traditional telecommunication technologies (e.g., high speed fiber optic data transmission, digital wireless communication technology, etc.).

As this is the first known foray into the use of Internet technology for cognitive evaluation, no prior research specific to the topic was available for comment and only two known articles have been published to date that combine Internet technology (specifically the WWW) with mental health service delivery (Lange et al., 2000; Tate et al., 2001). Dutch researchers, Lange et al., describe the use of an Internet-mediated protocol for the treatment of psychological dysfunction secondary to trauma. In their investigation, participants were assessed for psychological dysfunction on-line via

questionnaires before beginning a 5-week treatment program. On-line assessment of treatment goals was assessed immediately after treatment and after a 6-week follow-up period. After treatment, participants had improved significantly in the reduction of post-traumatic stress symptom reports and in overall endorsement of psychological well-being; treatment gains that were observed at the 6-week follow-up. While the Lange et al. results appear encouraging, conclusions about the relative effectiveness of the on-line treatment program stem from their use of a non-community based sample (i.e., Lange et al. recruited 36 undergraduate students for participation) and the lack of a control group comparison. A more recent and better-controlled study of the use of the Internet in a behavioral medicine application was performed by Tate et al. In this study, participants were randomly assigned to a 6-month weight loss program of either Internet-based didactics or Internet-based behavioral therapy. All participants were given face-to-face contact during one group weight loss session and access to a WWW site with organized links to WWW weight loss resources. Those participants enrolled in the behavioral therapy group received additional behavioral procedures, including a sequence of 24 weekly behavioral lesions via email, weekly online self-monitoring diaries with therapist feedback, and access to an online bulletin board to facilitate discussion among the group participants. Results from the Tate et al. study revealed that the Internet behavioral therapy group loss on average approximately 2 pounds more than the Internet education group at 6 months.

While no research to date has investigated the use of the Internet for remote cognitive assessment, research by Browndyke, Gouvier, & Waters (1999) and Browndyke, Santa Maria, Pinkston, & Gouvier (1998) was helpful in establishing the logistical process necessary for RNA. The Browndyke et al. studies, while limited to the remote administration and collection of questionnaire data, reported the effective use of the Internet for the collection of patient data related to mild traumatic brain injury and post-concussional symptom complaints. The current investigation was viewed as an extension of these studies with a shift in focus from the collection of questionnaire data to the administration and collection of cognitive evaluation data.

In order to demonstrate the utility of cognitive evaluation via RNA, the selection and creation of a prototype RNA measure was necessary. Due to the relative resistance to apparatus changes and ease of portability to computerization (Choca, Laatsch, Wetzel, & Agresti, 1997), the CT was chosen as the measure by which the RNA model would be empirically tested and validated.

Historical and Critical Review of the Category Test

Towards the beginning of his career, Ward Halstead, a professor of Physiology at University of Chicago, became interested in measuring the ability of brain-injured individuals to intuit commonalities among objects. Using sorting techniques with various three-dimensional objects, Halstead found that he was able to differentiate normal examinees from those who had suffered

brain damage based solely upon their categorization ability (Halstead, 1940). The large number of three-dimensional stimuli and scoring complexities of his original sorting procedure were abandoned for the development of the Halstead Category Test (HCT). Halstead substituted the objects from his original procedure for printed geometric figures, which were presented serially on a specially constructed apparatus. The original HCT stimuli were prepared on strips of white cloth attached to the circumference of a metal drum, which was placed behind a board with a viewing window. An electrical escapement was employed that was connected to a board with four response keys. Pressing the correct response key for test item would allow the escapement to move the drum forward to the next test stimuli, but pressing any of the three incorrect response keys for an item would not engage the escapement until the proper key was pressed. HCT response feedback was predicated upon the whether the examinee's initial response to a stimuli item was associated with an advancement (i.e., correct response) or non-advancement (i.e., incorrect response) to the next stimuli. This first version of the HCT consisted of 360 items, arranged into nine different subtests. Each of the nine subtests had a single organizing principle (e.g., Roman numerals, number of objects, oddity, quadrant, part/whole, and recognition), which ran throughout the subtest. Halstead told subjects to discover the organizing principle for each of the subtests based upon the correctness or incorrectness of their responses. A total test taking time limit was imposed, but during that time, subjects were

allowed to continue unaided in their solutions. The score from the original HCT was composed of the total number of errors made on the first response to the test stimuli from the nine subtests (Halstead & Settlage, 1943).

By 1947, when Halstead published his seminal work Brain and Intelligence (Halstead, 1947), the HCT apparatus and stimuli had evolved to a form very similar to a CT version widely used today – the Halstead-Reitan Category Test (HRCT; Reitan, 1969; Reitan & Wolfson, 1985), one of the oldest and most widely used neuropsychological assessment measure (Parsons, 1986) and viewed as a sensitive indicator of the brain dysfunction (Reitan & Wolfson, 1995). Rather than cloth-backed stimuli being unrolled from a drum, the new HCT apparatus employed a projector that sequentially presented the test stimuli on a screen facing the test subject. The revised HCT viewing screen was housed in a box approximately 22 x 14 x 16 inches and directly underneath the screen was a switchbox with four keys, numbered 1 to 4. The examiner for the revised HCT controlled the presentation of each stimuli frame and preset a "chime" and "buzzer" response feedback to the appropriate keys by means of a separate control box. A "chime" indicated that an examinee responded to the HCT stimuli correctly; whereas, a "buzzer" denoted an incorrect stimuli response. The change in HCT apparatus from an advance/non-advance response feedback system to auditory feedback significantly altered the course of the CT administration procedure. An early instruction manual for the revised HCT (Halstead & White, 1950) reveals that the change in response feedback

allowed for only one response per item (i.e., stimuli would advance to the next item irrespective of the correctness or incorrectness of a subject's response), which removed the ability for response correction for individual stimuli, and subsequently, added an element of short-term task retention to the CT assessment paradigm. The total number of items was reduced in the revised HCT to 208, the 9 original subtests were shortened to 7, and the subtest principles of roman numeral, number of objects, oddity, quadrant, completion, and recognition were retained. The revised HCT subtest structure, one that is found in all non-abbreviated CTs, is as follows: Subtest I is composed of eight Roman numeral stimuli items (numerals I – 4); Subtest II has twenty items of various linearly arranged geometric shapes (e.g., squares, vertical lines, circles, etc.); Subtest III was shortened to forty stimuli items also composed of linearly arranged geometric figures, but differing from the preceding subtests in the addition of color to some of the stimuli; Subtest IV departs from the linear stimuli seen in Subtests II and III by presenting forty stimuli items that are drawn based upon a Cartesian grid system (i.e., equidistant vertical and horizontal division of various geometric figures); Subtest V also adopts the Cartesian division of stimuli for more than half of the forty items comprising this subtest, while the remaining stimuli from Subtest V are composed of partially completed figures represented by solid and dotted vertical or horizontal lines; the forty stimuli in subtest VI continue the pattern of solid or dotted partially completed figures noted in Subtest V; and lastly, Subtest VII is composed of an

amalgamation of twenty stimuli from Subtests II – VI and was conceptualized by Halstead as a task recognition component (Halstead, 1947; Halstead & White, 1950).

Based upon the investigation of non-abbreviated CT subtest error score data, the factor structure of subtest scores appears to include three factors, one of which appears to reflect task memory (Boyle, 1988; Fischer & Dean, 1990). Johnstone, Holland, and Hewett (1997) described three factors: Symbol Recognition/Counting, comprised of errors from Subtests I and II; Spatial Positional Reasoning, derived from Subtest III, IV, and VII errors; and Proportional Reasoning, composed of errors from Subtests V and VI. That is, the CT appears to assess the ability to count, to identify the correct element in a spatial array based on learning a principle involving oddity or location, and to abstract the principle of proportion regardless of variations in the form or number of elements in the stimulus array. The factors of spatial positional reasoning and proportional reasoning appear consistently in both normal and patient populations (Fischer & Dean, 1990; Kelly, Kundert, & Dean, 1992; Livingston, Gray, & Haak, 1996); whereas, some researchers have failed to detect a symbol recognition/counting factor based upon a general lack of error scores on CT subtests I and II (Allen, Goldstein, & Mariano, 1999). Four factors, labeled "central integrative, abstraction, power, and directionality," were found when the revised HCT was analyzed with other tests from the original Halstead Neuropsychological Assessment Battery (Choca et al., 1997;

Halstead, 1947). When analyzed with contemporary neuropsychological assessment instruments, the HRCT was found to load on factors of general intelligence (Barnes & Lucas, 1974; Boyle, 1988; Holland & Wadsworth, 1976), complex spatial reasoning (Aftanas & Royce, 1969; Lansdell & Donnelly, 1977; Russell, 1974), and "fluid" intellectual abilities (Cullum, Steinman, & Bigler, 1984).

The total number of errors has been the measure traditionally extracted from the CT. To assess the error score effectively, the use of norms and standardized score conversions have been adopted for most of the non-abbreviated CT versions (e.g., HRCT; Heaton, Grant, & Matthews, 1991), while most short forms of the CT use regression equations to predict the total error score (Caslyn, O'Leary, & Chaney, 1980). In addition to a CT total error score, some studies have investigated examinee response patterns (Brandt & Doyle, 1983; Simmel & Counts, 1957), item analysis (Laatsch & Choca, 1991), and the use of reaction times (Beaumont, 1975; Rattan, Dean, & Fischer, 1986).

The first systematic investigation of CT response patterns was undertaken exhaustively by Simmel & Counts (1957), who argued that four factors co-determine an examinee's response choice characteristics: (1) the perceptual characteristics of stimulus configuration; (2) the application of new items to a previously learned subtest principle; (3) an "einstellung effect," defined as the priming of a counting concept by the CT apparatus configuration (i.e., numbered response keys); and (4) idiosyncratic response tendencies. In

addition to CT response choice characteristics, Simmel & Counts made an important distinction between "essentially" and "incidentally" correct responding. An "essentially" correct examinee response is based upon their application of the learned subtest principle; whereas, "incidentally" correct responses are determined by some prominent feature of the CT stimulus constellation (or by the immediately preceding subtest principle), which happens to be correct – basically, an accidental correct response. Bearing these characteristics and response possibilities in mind, Simmel & Counts make a strong argument that factors other than the application or non-application of a subtest principle may determine an examinee's test outcome, and, in turn, question the purity of the category learning process in the CT paradigm. Regardless of the difficulties with CT characteristics, Simmel & Counts concede that even in their study of 26 normal controls and 35 anterior temporal lobectomy patients, the CT was very effective in distinguishing normals from brain injured patients. They postulate that, in addition to difficulties in categorization abilities, brain injured examinee's are negatively affected by the complexities of "essentially" and "incidentally" correct stimuli items and cannot cope effectively with the myriad of CT response factors. Brandt & Doyle (1983) also looked beyond the total CT error score and examined the response pattern of adolescent drug users in an effort to investigate possible difficulties with tracking and set shifting abilities in this patient population, and other research efforts have uncovered several different response patterns that are

demonstrably independent of the total error score (Laatsch & Choda, 1991, 1994).

An item analysis of the HRCT by Laatsch and Choca (1991) uncovered 45 dysfunctional items. Items on Subtests I and II were found to be too easy to yield useful information, even if they serve to familiarize the examinee with the mechanics of the test. Laatsch and Choca demonstrated an uneven progression of the mean item difficulty in successive subtests, with an especially abrupt jump in the difficulty levels of Subtests I and II to those of Subtest III (Choca et al., 1997). They suggest that this jump in difficulty may pose a problem for some patient populations independent of the intended measurement of the CT; an argument similar to that made by Simmel and Counts (1957).

Using a computer version of the CT, Choca and colleagues (Choca, Laatsch, Garside, & Arnemann, 1994) reported an average reaction time of 8.1 seconds per item response. Extrapolated from their results, the average total test taking time for the Choca et al. computerized version of the HRCT was approximately 28 minutes. Prior computerized reaction time research by Beaumont (1975) indicated an average response latency of 5 seconds per item, which would lead to an average total test taking time of 17 minutes. Rattan et al. (1986) proposed that average reaction time be considered in evaluating CT performance. Rattan and colleagues suggested that the addition of a timing factor to the CT assessment paradigm might prove a useful indicator of learning

efficiency. Lastly, Choca et al. (1997) suggest that a poor CT error score combined with very fast average reaction times usually betrays impulsivity and lack of motivation on the part of the examinee; a finding which indicates that the addition of a response timing factor to the CT may prove to be a useful measure of task effort or malingering.

The revised HCT stimuli, number of subtests, and response rules remain virtually identical to those currently employed in the HRCT and other non-abbreviated CTs [e.g., Booklet Category Test (BCT; DeFilippis, McCampbell, & Rogers, 1979; DeFilippis & McCampbell, 1991), Remote Neuropsychological Assessment – Category Test (RNA-CT; see Appendix B)]. Non-abbreviated versions of the CT are also available in paper-and-pencil versions (Adams & Trenton, 1981; Wood & Strider, 1980), a portable version (Slay, 1984), and a card version (Kimura, 1981). An Intermediate version of the revised HCT was constructed by Reitan and colleagues (ICT; Reed, Reitan, & Klove, 1965), which is meant to be administered to children between the ages of 9 to 14, and a version of the CT for children under the age of nine is available, as well (Boll, 1993). Many short forms of the CT have been constructed over the years, but will not be addressed specifically in this review (see Boyle, 1986; Caslyn et al., 1980; Moehle, Fitzhugh-Bell, Engleman, & Hennon, 1988; Russell & Levy, 1987; Wood & Strider, 1980; Wetzel & Boll, 1987, for examples).

Objections have been raised regarding the use of any version of the CT other than the HRCT. Reitan and Wolfson (1985) challenged the BCT on the

grounds that the administration procedure alters the inherent properties of the CT (i.e., booklet vs. projector/viewer apparatus), which negatively affects the validity and reliability of the procedure. The time to complete the non-abbreviated CT versions can be considerable for some patients, but most individuals complete the test in less than 40 minutes (Finlayson, Sullivan, & Alfano, 1986). For some, this amount of time may seem excessive when it is invested in obtaining the score from just one measure; a factor that has fueled the creation of many CT short forms. As with Reitan and Wolfson's (1985) concerns, Vanderploeg and Logan (1989) contend that the reduction in number of the conceptual shifts tested by any of the CT short forms will impair the test's validity. Russell and colleagues (Russell & Barron, 1989; Russell & Levy, 1987) counter these concerns by pointing out that statements about apparatus or short-form CT alterations negatively affecting test validity have not been supported by empirical data. To underscore Russell and colleagues' point, almost all of the non-abbreviated and short form CT versions that have been studied to date have been found to be reasonably equivalent to the HRCT (e.g., Adams & Trenton, 1981; Berger, Chibnall, & Gfeller, 1997; Boyle, 1986; Byrd & Ingram, 1988; Caslyn et al., 1980; Choda & Morris, 1992; DeFilippis, & McCampbell, 1991; DeFilippis et al., 1979; Gregory, Paul, & Morrison, 1979; Holtz, Gearhart, & Watson, 1996; Kimura, 1981; MacInnes, Forch, & Golden, 1981; McCampbell & DeFilippis, 1979; Russell & Levy, 1987; Slay, 1984; Wetzel & Boll, 1987). Performance on the CT does not appear to be affected by the

omission of the usual instructions (i.e., revised HCT or HRCT instructions) regarding the changes in the underlying principle from one subtest to another (Rothke, 1986), and visual, rather than auditory response feedback, with hearing impaired examinees has been successfully used to administer the HRCT (Kelly, 1995; Wood & Strider, 1980). Thus, from the number of studies listed above, it would seem that the issue of whether the examinee can perform the CT is predominant enough that the effects of variation in administration, or in the manner of presentation, appear to be minor or irrelevant (Choca et al., 1997).

Upon investigating individual differences between examinees, it is very clear that age is a primary factor moderating performance on the CT (Elias, Robbins, Walter, & Schultz, 1993; Ernst, 1987; Fromm-Auch & Yeudall, 1983; Heaton, Grant, & Matthews, 1986; Mack & Carlson, 1978; Prigatano & Parsons, 1976; Query, 1979; Reed & Reitan, 1963; Reitan & Davidson, 1974; Reitan & Wolfson, 1985). The strongest relationship between age and CT performance was cited by Leckliter and Matarazzo (1989), who reported a correlation coefficient of .54 between age and total CT errors in three normative samples (Choca et al. 1997).

Educational level has also been known to factor in the prediction of CT total error scores (Finlayson, Johnson, & Reitan, 1977; Golden, Osmon, Moses, & Berg, 1981; Heaton et al., 1986; Prigatano & Parsons, 1976; Reitan & Wolfson, 1985). Leckliter and Matarazzo (1989) also calculated relatively a

relatively strong inverse relationship (i.e., r = -.31) between educational obtainment and CT total errors in their normative studies (Choca et al., 1997).

There appears to be an interaction between the age and educational level variables. Heaton et al. (1986) point out that up to age 60, less educated HRCT examinees show greater age-related impairments than their more educated same-age cohorts. After the age of 60, however, all examinees in their normative database tended to perform poorly in equal measure on the HRCT, regardless of their educational status (Choca et al., 1997). Several possible explanations have been posited to account for this homogenization of education-related contributions in the elderly. In attempting to answer this observed effect in CT performance, Choca et al. (1997) suggest the possibilities that "poorly educated individuals have a greater proclivity toward early brain damage. Or perhaps the better educated enjoy a 'brain reserve' that allows them to keep functioning at a higher level, in spite of age-related losses in brain efficiency" (p. 64).

Kupke (1983) detected subtle gender differences in HRCT performance, but to date there has been no convincing evidence that gender bias plays a significant role in moderating CT outcome (Elias et al., 1993). However, cultural effects involving gender have been reported by Cuevas and Osterich (1990). In a study of cross-cultural effects on BCT performance, Cuevas and Osterich demonstrated that European women tended to obtain a higher number of BCT errors relative to European male counterparts and American women.

Additional cultural effects were detected by Arnold, Montgomery, Castaneda, & Longoria (1994), who observed that Mexican and Mexican-Americans performed less well on the HRCT than an Anglo American control group.

For all of the possible individual differences that have been shown to moderate CT performance, it is important to recognize that most of these relationships tend to dissipate when the CT is studied in clinical populations (Leckliter & Matarazzo, 1989; Prigatano & Parsons, 1976; Query, 1979; Reitan & Davidson, 1974; Russell, 1997; Vega & Parsons, 1967). Leckliter and Matarazzo (1989) suggest that brain damage, or the emotional interference present in a psychiatric populations, has an overpowering effect on CT performance, reducing the test performance variability afforded to the aforementioned demographic contributions.

HRCT test-retest reliability has been shown to be low (r = .60) in a normal population, presumably due to the benefit obtained from task familiarity (Matarazzo, Weins, Matarazzo, & Goldstein, 1974). Russell (1992), however, in a reinterpretation of the Matarazzo et al. (1974) test-retest reliability data found that by expanding the range of CT scores, he was able to obtain an improved test-retest coefficient in a normal population. Matarazzo, Matarazzo, Wiens, Gallo, and Klonoff (1976) found that the HRCT test-retest reliability in patient populations tended to increase as an examinee's performance worsened. The relatively poor test-retest reliability in normals and increasing test-retest reliability in patient populations reported by Matarazzo et al. has been

supported by other investigators and suggests that for some patients CT performance is relatively invariant over time (Bornstein, Baker, & Douglass, 1987; Dodrill & Troupin, 1975; Eckhardt & Matarazzo, 1981). This point is bolstered by the finding that when CT learning is defined as the reduction of errors over time during the first administration of the test, there is only a modest contribution to the outcome of the final CT error score (Bertram, Abeles, & Snyder, 1990). A split-half reliability of .90 or above has been reported consistently in the research literature (Moses, 1985; Shaw, 1966).

Research investigations examining the relationship of the CT to other instruments have generally led to only modest correlations (Choca et al., 1997). Of the other measures comprising the HRNAB (Reitan, 1969; Reitan & Wolfson, 1985), the CT demonstrated some relationship with the total performance time of the Tactual Performance Test and Form B of the Trail Making Test in a psychiatric patient population (Goldstein & Shelly, 1972). The correlation between the HRCT total error score and the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1981) intellectual quotient (IQ) has been found to range between r_s of -.30 and -.78 (Barnes & Lucas, 1974; Beaumont, 1975; Cullum et al., 1984; Goldstein & Shelly, 1972; Landsell & Donnelly, 1977; Reitan, 1955; Reitan & Wolfson, 1985), with more significant correlative relationships being found between HRCT total error score and WAIS Performance IQ (Corrigan, Agresti, & Hinkeldey, 1987; Goldstein & Shelly, 1972; Landsell & Donnelly, 1977; Weins & Matarazzo, 1977). Among the Wechsler-Bellevue subtests, the

highest correlations reported have been between the HRCT total error score, Digit Symbol and Block Design subtests (Reitan, 1956).

The test that many clinicians and researchers see as the most similar to the CT is the Wisconsin Card Sorting Test (WCST). However, even with this instrument, only modest correlations have been reported with the BCT and HRCT (Donders & Kirsch, 1991; Pendleton & Heaton, 1982; Perrine, 1993). Perrine (1993) suggests that the modest relationship is due to the fact that the two tests are measuring different aspects of concept formation; a hypothesis that appears to be supported by the functional neuroimaging research of Adams et al. (1995).

Halstead (1947) noted that patients with frontal lobe brain injury did more poorly on the CT than other brain injured patient samples; a localizing finding which has not been consistently replicated by other researchers (Bornstein, 1986; Choda et al., 1997; Golden et al., 1981; Reitan & Wolfson, 1995; Shure & Halstead, 1958). This inability to consistently replicate Halstead's (1947) results has not, however, doomed the clinical utility of the CT, for even though it is not seen as a useful measure in the differentiation of frontal lobe brain injury from different brain injury types, it does differentiate, with a high degree of sensitivity, generalized brain dysfunction from normal brain function (Bornstein, 1986; Choca et al., 1997; Golden et al., 1981; Matthews, Shaw & Klove, 1966; Reitan, 1969; Reitan & Davidson, 1974; Reitan & Wolfson, 1995; Watson, Thomas, Anderson, & Felling, 1968). However, false positive rates of

as much as 18% have been reported for the BCT and HRCT (DeFilippis et al., 1979; Reitan, 1955), indicating that a high number of errors on the CT can result from a variety of problems, not necessary indicative of brain dysfunction. Shute and Huertas (1990) conceptualized the CT as a measure of Piaget's formal operations stage – the most advanced stage of cognitive development, characterized by effective reasoning and problem-solving capacities. According to Shute and Huertas, this Piagetian stage is only fully reached by approximately half of neurologically intact adults, and conceptualizing the CT in this manner may help explain the variability of CT error scores and false positive rates found in prior research endeavors. In summary, good scores on any version of the CT can be assumed to reflect an intact brain, reasonable intellectual abilities, maturity in cognitive development, and capacity to think with concentration and efficiency. However, poor scores could have various and multiple determinants, including brain damage, low level of intellectual ability, primitive cognitive development, or emotional interference with the capacity to use cognitive resources (Choca et al., 1997).

The CT appears to be quite adaptable to computer administration. In conventional modes of CT presentation (e.g., HRCT or BCT), either a slide projector presents the test stimuli on a monitor, an examinee manipulates a series of levers to indicate answer choice, and a bell or buzzer indicates his/her correct or incorrect responses, or the examinee is presented with stimuli in a folio, asked to point to answer choices on a numbered strip of paper, and

provided with verbal feedback indicating correct or incorrect responses. These procedures are similar to what would occur during a computerized administration of the CT and have proven rather simple to replicate using modern computer hardware and software. Various computerized versions of the CT have been created that provide feedback, change images, score examinee test data, and collect response times (Beaumont, 1975; Choca, 1987; Miller, 1993), and an adaptive computerized short form of the CT has also been programmed, which administers items based upon on examinee's performance (Choca et al., 1994; Laatsch & Choca, 1994). Prior research appears to suggest that the CT assessment paradigm may be robust to instrumentation changes (Russell & Barron, 1989; Russell & Levy, 1987), but the literature has been somewhat equivocal regarding various computerized versions (Beaumont, 1975; Berger, Chibnall, & Gfeller, 1994, 1997).

The first application of computer technology in the administration of the CT was undertaken by Beaumont (1975), who used a LINC-8 computer to initiate CT stimuli change and record and monitor responses. In addition to being the first to automate the CT using computers, Beaumont was also the first investigator to systematically record response latency variables during his task. The results derived from his study were, unfortunately, not as groundbreaking and led to the conclusion that the computerized CT was not a valid substitute for the standard version. Beaumont based his conclusions on the outcome of a brain damage and patient control sample comparison, which

indicated that when the criterion of 50 or more total errors on the computerized CT was employed as an indicator of brain damage (Reitan & Wolfson, 1985), 30% of the brain damaged group and 70% of the patient control group were misclassified. A direct comparison of the equivalency between Beaumont's computerized CT and the HRCT was not performed, severely limiting any conclusions which can be drawn from his study regarding the possible effects of apparatus changes contributing to the poor CT outcome. The application of computers in CT administration would remain dormant until the late 1980s, at which point, Choca constructed the next iteration of the computerized CT (Choca, 1987). A vast improvement over the attempt by Beaumont, the Choca computerized CT had the advantages of microcomputer technological advances, which fully automated the CT with the exception of verbal instructions and prompts necessary for task completion. In addition, the Choca computerized CT demonstrated an acceptable level of equivalence with the HRCT in an inpatient veteran sample (Choca & Morris, 1992). In contrast to the Choca and Morris results, Berger et al. (1994) found that the use of the Choca computerized CT in a private clinic patient sample resulted in a significantly greater number of errors than the administration of the HRCT, which placed the construct validity of the Choca computerized CT into question. Unfortunately, Berger et al. failed to randomize their study groups and did not administer the Choca computerized CT properly (i.e., observation of computerized test administration to lend assistance with task or apparatus as needed), which

called into question the veracity of their results. A more recent version of the Choca computerized CT has been developed, which adapts the presentation of CT stimuli based upon the performance of an examinee (Choca et al., 1994). Response timing variables can also be collected with the Choca et al. version which may assist in the clinical interpretation of the test (Laatsch & Choca, 1994). The last known computerized CT reported in the literature was developed by Miller (1993), who adapted the CT for use with the Macintosh computer operating system. In a comparison of the Miller computerized CT with the BCT and HRCT in brain-injured and normal control samples, Mercer, Harrell, Miller, Childs, and Rockers (1997) were able to demonstrate that the three versions of the CT were robust to instrumentation changes. No significant differences were detected between groups on the CT total error variable and a lack of interaction between injury group and test version suggested that the measures were essentially equivalent across samples.

In summary, the CT has been a time tested neuropsychological assessment paradigm tapping abstract conceptualization skills, and while numerous versions of the test have been developed, the core features of sensitivity to brain dysfunction, emphasis on spatial positional and proportional reasoning abilities, and the relative ease of administration and scoring persist in practically all of its iterations.

PURPOSE OF THE STUDY

Most of the problems noted in prior computerized CT research (Beaumont, 1975; Berger et al., 1994, 1997) have been viewed as being due, in part, to standardization difficulties generated by the use of varied computer instrumentation and apparatus (i.e., line printer computer, keyboard vs. mouse, operating system and monitor differences, etc.), poor methodological control, task confusion (i.e., examiner did not stay in the room during computerized CT administration), and the variable approximation to conventional HRCT administration guidelines (Choca et al., 1997). In order to address some of these concerns and to lay the groundwork for the application of neuropsychological assessment via remote service delivery (i.e., the RNA model), a new computerized, Internet-based CT was created. The following investigation is an empirical examination of the RNA-CT, which differs markedly from prior computerized CT versions. The RNA-CT utilizes multimedia techniques, to more closely approximate conventional testing conditions, and a WWW browser delivery system to circumvent instrumentation difficulties (i.e., RNA-CT is independent of operating system type) experienced by prior computerized CT versions. By closely simulating conventional CT administration procedures and careful control of extraneous testing variables, it was hypothesized that the RNA-CT would demonstrate equivalency with a conventional version of the CT assessment paradigm (e.g., BCT) in a normal subject sample. Equivalence comparisons of CT total error score central

tendency and subtest error score factor structures were carried out to investigate the convergent and construct validity of the RNA-CT using the BCT as an established comparison measure. In addition to the primary equivalence comparisons of total error and subtest error factor structures, the RNA-CT and BCT groups were compared on stimuli response timing variables (e.g., total test time and time per subtest); a relatively untapped property of the CT assessment paradigm. It was hypothesized that, like the primary equivalence comparisons of total error and subtest error factor structure, the RNA-CT and BCT groups would not differ significantly on any of the response timing variables.

Test equivalence methodologies and task construction were planned in accordance with the APA Testing Standards (1986), standards referred to by Hofer (1985) and stipulated by the APA Division 40 Task Force on Computer-Assisted Neuropsychological Evaluation (American Psychological Association Committee on Professional Standards & Committee on Psychological Tests and Assessment, 1987).

METHODOLOGY

Subjects

Subjects for the current research investigation were recruited from the Louisiana State University – Psychology Department undergraduate subject pool. Those subjects who volunteered to participate in the study received psychology course extra credit points equivalent to 1.5 hours; the total estimated time for experiment completion. A power analysis indicated a minimum of 34 participants per group was needed to ensure sufficient power for a medium effect size (i.e., $\underline{d}$ = .60) at a power level of .80 and a significance level of .10 (Cohen, 1992). To address any possible subject attrition difficulties or the post-hoc discovery of exclusion criteria being met for any of the participants, 84 subjects were recruited for study participation. As subjects arrived for the experiment in pairs of two, they were randomized by coin toss to one of two testing conditions; the RNA test administration (i.e., RNA-CT) or the manual test administration (i.e., BCT). Seventy-one subjects were retained for the project data analyses – 37 subjects in the RNA-CT group and 34 subjects in the BCT group. Sixteen of the 84 participating subjects were excluded from the data analyses based upon the presence of exclusionary criteria in their remote or recent history (e.g., significant head injury, seizure disorder, psychiatric difficulties, substance abuse, etc.).

Exclusionary criteria were implemented to prevent the study groups from differing on variables of known or suspected significance to CT outcome (see

Table 1. Screening Criteria

Variables	Included for Analysis	Excluded from Study
Age	17 – 35	< 17 or > 35
Estimated IQ[a]	86 – 114	≤ 85 or ≥ 115
Neurological/Psychiatric Illness	No history of significant neurological and/or psychiatric illness	Evidence of current or prior neurological and/or psychiatric illness
Computer-related Anxiety[b]	No	Yes

[a]Estimated WAIS-R full-scale IQ derived from the Shipley Institute of Living Scale (SILS; Shipley, 1940; Zachary, Crumpton, & Spiegel, 1985). [b]Defined as a score of > 1.5 s.d. above the normative mean on the Computer Anxiety Rating Scale (Rosen et al., 1992).

Table 1). An age criterion was implemented to minimize the variance in the CT error score that could be accounted for by age-related factors; a variable that has been cited as mediating CT error rate, particularly in individuals over 50 (Boyle, 1986; Choda et al., 1997; Heaton et al., 1991; Mack & Carlson, 1978; Prigatano & Parsons, 1976; Query, 1979; Reitan & Wolfson, 1995). To address this issue of age-related mediation of CT error rates, subjects in the current investigation were restricted to individuals between the ages of 17 and 35 years. Exclusionary criteria controlling for the possible effect of intellectual functioning on CT performance were also employed, as many researchers have noted a relationship between general intellectual ability and CT error rates (Choda et al., 1997; Finlayson et al., 1977; Heaton et al., 1991; Lansdell & Donnelly, 1977). In an attempt to control for intellectual ability differences

between groups, subjects were administered the Shipley Institute of Living Scale (Shipley, 1940); a brief, pencil and paper measure of verbal ability and logic that correlates well (e.g., rs from .73 to .87) with full-scale IQ scores from the Wechsler Adult Intelligence Scale, Revised (WAIS-R; Wechsler, 1981; Zachary et al., 1985). Participants were restricted to those subjects whose Shipley estimated full-scale IQ was within one standard deviation above or below the "average" WAIS-R IQ range (i.e., est. WAIS-R IQ $\leq$ 114 and $\geq$ 86). The sensitivity of CT outcome to various neurological and psychiatric conditions is well known and has been extensively documented (Anderson, 1994; Bornstein, 1986; Choda et al., 1997; DeFilippis et al., 1979; Mercer et al., 1997; Reitan & Wolfson, 1985; Reitan & Wolfson, 1995; Shaw, 1966). As a result, efforts were taken to exclude subjects from the study with neurological or psychiatric conditions that could negatively affect test performance in either group (see Appendix A). To prevent the previously untested, but suspected, issue of computer-related anxiety from negatively affecting RNA-CT performance, the Computer Anxiety Rating Scale (CARS; Rosen et al., 1992) was administered to prospective subjects randomized to the RNA administration group. Those subjects whose CARS score suggested a level of computer-related anxiety greater than 1.5 s.d. above the normative mean score suggested in the CARS normative manual (i.e., CARS total $\geq$ 80) were excluded from the study.

No significant differences in controlled or uncontrolled (i.e., gender and handedness) demographic variables were detected between the study groups (see Table 2). However, significantly greater proportions of female ($p < .05$) and right-handed ($p < .001$) participants were noted within each group. The finding of gender bias was not wholly unexpected, as the study groups were drawn from the same population, which tends to be more heavily weighted towards female participants, and the significant within-group results for handedness only mimic the proportion of right- and left-handed individuals noted in the general population.

By restricting the groups on the aforementioned demographic variables, a level of experimental control was established, which allowed for the creation of

Table 2. Demographic Comparisons

Variables		RNA-CT[a]	BCT[b]	p-values
Age	Mean (SD)	20.86 (2.08)	21.12 (2.25)	.63
Education	Mean (SD)	14.62 (1.30)	14.50 (.99)	.66
Estimated IQ[c]	Mean (SD)	103.70 (6.81)	105.00 (7.10)	.44
Gender (M/F)[d]		9 / 28	12 / 22	.51 / .37
Handedness (R/L)[d]		34 / 3	30 / 4	.62 / .71

Note. Analyses were conducted using independent sample T-test comparisons, unless noted otherwise.
[a]$n = 37$. [b]$n = 34$. [c]Estimated WAIS-R full-scale IQ derived from the Shipley Institute of Living Scale (SILS; Shipley, 1940; Zachary et al., 1985). [d]Chi-square statistic (X^2).

relatively homogeneous study samples. As a result, the samples were thought to be minimally biased by factors other than those inherent to the main variable of interest (i.e., RNA administration vs. manual administration); thus, increasing the overall confidence that could be placed in the results of the equivalence comparisons.

Measures

Screening Questionnaire. A general screening questionnaire (see Appendix A) was used to obtain information about participant demographics (e.g., gender, age, education, etc.), as well as the following neurological and/or psychological conditions: 1) a history of head trauma greater than mild, uncomplicated severity or a history of repeated, uncomplicated mild severity head trauma; 2) a history of seizure or seizure disorder; 3) a history of central nervous system disease (e.g., infection, tumor, vascular, developmental, degenerative, toxic, metabolic, and demyelinating); 4) a history of stroke or transient ischemic attack; 5) exposure to electroconvulsive therapy or pharmacotherapy for psychiatric illness; 7) a history of psychiatric illness, including panic disorder, post-traumatic stress disorder, obsessive-compulsive disorder, major depression, dysthymia, mania, and psychosis; and 6) current excessive alcohol and/or drug use (e.g., alcohol, marijuana, cocaine, amphetamines, barbiturates, and hallucinogens). The screening questionnaire was administered by the experimenter in a brief interview format, which

allowed for the clarification and follow-up of endorsed exclusionary criteria. The screening questionnaire took, on average, approximately 10 minutes to complete.

Computer Anxiety Response Scale (CARS). Developed by Rosen et al. (1992), the CARS is composed of twenty questions ranging from general technology contact (e.g., "re-setting a digital clock after the electricity has been off") to varying levels of computer-specific experience (e.g., "learning to write computer programs") and is intended as a self-report measure of computer-related anxiety symptoms and cognition. Subject responses are anchored 5-point Likert ratings with a range from "not at all," indicating a low level of subjective anxiety, to "very much," indicating a high level of subjective anxiety. Once completed, the CARS responses are summed to yield a total score (minimum CARS score = 20; maximum CARS score = 100). Based upon the CARS standardization sample, a total score $\geq$ 80 (i.e., 1.5 s.d. above the normative mean CARS score) is indicative of a potentially problematic level of computer-related anxiety symptoms.

Shipley Institute of Living Scale (SILS). The Shipley Institute of Living Scale (Shipley, 1940) is a 2-page brief screening instrument often used to estimate current intellectual functioning. The SILS has been widely used in research and clinical settings where time may be limited, yet a gross estimation of intellectual skills is necessary for subject selection. The SILS is divided into two components, a verbal synonym knowledge subtest comprised of 40

multiple-choice items (e.g., "jocose = humorous, paltry, fervid, or plain") and 20 completion problems tapping logical abstraction and sequencing abilities (e.g., "AB BC CD D_"). Each SILS subtest is timed for 10 minutes and the sum of the number of items correctly completed within the time limits serves as the total score. Zachary et al. (1985) developed regression equations allowing for the conversion of total SILS scores to estimated WAIS-R full-scale IQ (FSIQ) scores, and they cited data from their conversion study which indicated that the estimated WAIS-R FSIQ and SILS scores correlate to a high degree (r = .87). Other researchers, however, indicate that SILS and WAIS-R FSIQ scores share a more modest positive correlation (r = .73; Dalton, Pederson, & McEntyre, 1987).

Booklet Category Test (BCT). The BCT (DeFilippis et al., 1979; DeFilippis & McCampbell, 1991) was developed as a portable version of the HRCT (Reitan, 1969; Reitan & Wolfson, 1985); the mechanical apparatus of which has been seen as cumbersome and expensive. Aside from minor revisions in task instructions and feedback modality (i.e., verbal feedback exchanged for the bell and buzzer feedback), the BCT is essentially the same measure as the HRCT (Byrd & Ingram, 1987; Choca et al., 1997; DeFilippis et al., 1979; DeFilippis & McCampbell, 1991; MacInnes et al., 1981). The BCT is composed of near exact replications of the 208 stimulus items from the revised HCT and HRCT. The stimuli are divided into seven separate subtests of unequal length and are sequentially presented in a binder format with each stimuli centered on 8.5" x

10" card stock paper (see Apparatus Section for more detail). Like the revised HCT and HRCT, the stimuli in each BCT subtest conform to an underlying rule or principle (e.g., total number of items, ordinal placement of a "odd" item in an array, etc.), which persists throughout that subtest. The rule or principle of a subtest may or may not carry over into a subsequent subtest. Individuals taking the BCT are asked to attend to a subtest stimulus figure, then to determine a number between one and four that may be represented by the figure subtest principle or rule. Verbal feedback from the test administrator, as to whether a subject's choices are correct or incorrect, allows the participant to change strategy and intuit from experience the underlying idea or principle unifying subtest stimuli. As with the HRCT, the BCT yields a total error score out of 208 responses, and although it is an un-timed test, the estimated completion time typically ranges from 20 to 40 minutes (Choda et al., 1997; Finlayson et al., 1986).[1]

DeFilippis et al. (1979) in their development of the BCT, compared performance on the HRCT and BCT in large samples of normal and alcoholic individuals. Within-subject comparisons for both groups on the HRCT and BCT yielded correlations of .91 for normals and .89 for alcoholics, suggesting a high-degree of equivalence between the two measures in normal and patient populations. Further data from DeFilippis & McCampbell (1991) suggest that the BCT has a robust level of discrimination between normal and brain injured

[1]Interested readers are referred to the literature review section for more information on the CT subtest principles employed by the BCT.

groups, similar to that cited for the HRCT (Bornstein, 1986; Choca et al., 1997; Doehring & Reitan, 1962; Holtz et al., 1996; Parsons, Jones, & Vega, 1971; Reitan, 1955; Reitan & Wolfson, 1995).

Remote Neuropsychological Assessment – Category Test. Given the CT's relative response simplicity, strong diagnostic utility, and wide-spread use, the CT assessment paradigm was a logical choice for the construction of a prototype computerized, Internet-based, neuropsychological assessment measure – the Remote Neuropsychological Assessment – Category Test (RNA-CT; see Appendix B). The RNA-CT was constructed to utilize item response feedback similar to the HRCT (i.e., bell and buzzer), combined with the addition of visual cues (e.g., green and red lights) that add an additional mode of response feedback for the examinee. The instructions for the RNA-CT are similar to those used by the BCT and HRCT and reflect a combination of instructional material from both tests, as well as instructions from the revised HCT. The alterations made to the test instructions were limited to those necessitated by the computerization of the test, and are especially apparent when referring to the method of subject responding (e.g., "...click on the numbered buttons on the screen with your mouse," instead of, "point to the number on the (paper) strip," indicated by the BCT instructions.). However, the RNA-CT does differ markedly from the HRCT and BCT in method of instruction presentation. Rather than instructions for task completion being only read to the test participant, the RNA-CT instructions are presented in text

form on the computer monitor and simultaneously in auditory form by the computer speakers. This dual coding of instructions and response feedback (i.e., visual and auditory) was added to the RNA-CT to allow for clear task comprehension and to expand the CT assessment paradigm for possible use with the hearing-impaired; a population disadvantaged by the auditory-dependent instruction and feedback of all of the other CT versions. The subtest composition, stimuli, and categorization principles are virtually identical to those initially developed by Halstead and used in the BCT and HRCT (see Literature Review section), and the scoring of the RNA-CT follows the same conventions as the HRCT and BCT (i.e., total number of errors out of 208 stimulus items). In addition to a total error score, the RNA-CT provides a method of determining the number of errors per subtest, as well as the total and average stimuli response time per subtest measured in milliseconds (ms).

Apparatus

The BCT (DeFilippis & McCampbell, 1991) materials included two large 9" x 12" three-ring binders, each containing test stimuli on 8" x 10" heavy stock paper, and a 4" x 2" strip of heavy stock paper listing the response stimuli (i.e., Arabic numerals 1 – 4). The BCT stimuli are approximately 4" in height and vary in width from 3" to 10".

Administration of the RNA-CT was performed on a Dell Pentium 166MHz personal desktop computer (model type, OptiPlex GM+5100), equipped with 16 MB of RAM, 32-bit file system/virtual memory, and 100MB of hard disk storage

space. Display of RNA-CT stimuli was carried out on a 15" viewable size monitor using an S3 SVGA graphics card, and the graphical resolution for the computer was set to a 256-color palette. RNA-CT data transmission was carried out over a U.S. Robotics 56-Kbit modem using an encrypted Internet connection between the laboratory computer and the remote server. Access to the server was limited via secure socket layer (SSL) technology, which encrypted all data transfers from the server to the laboratory computer (see

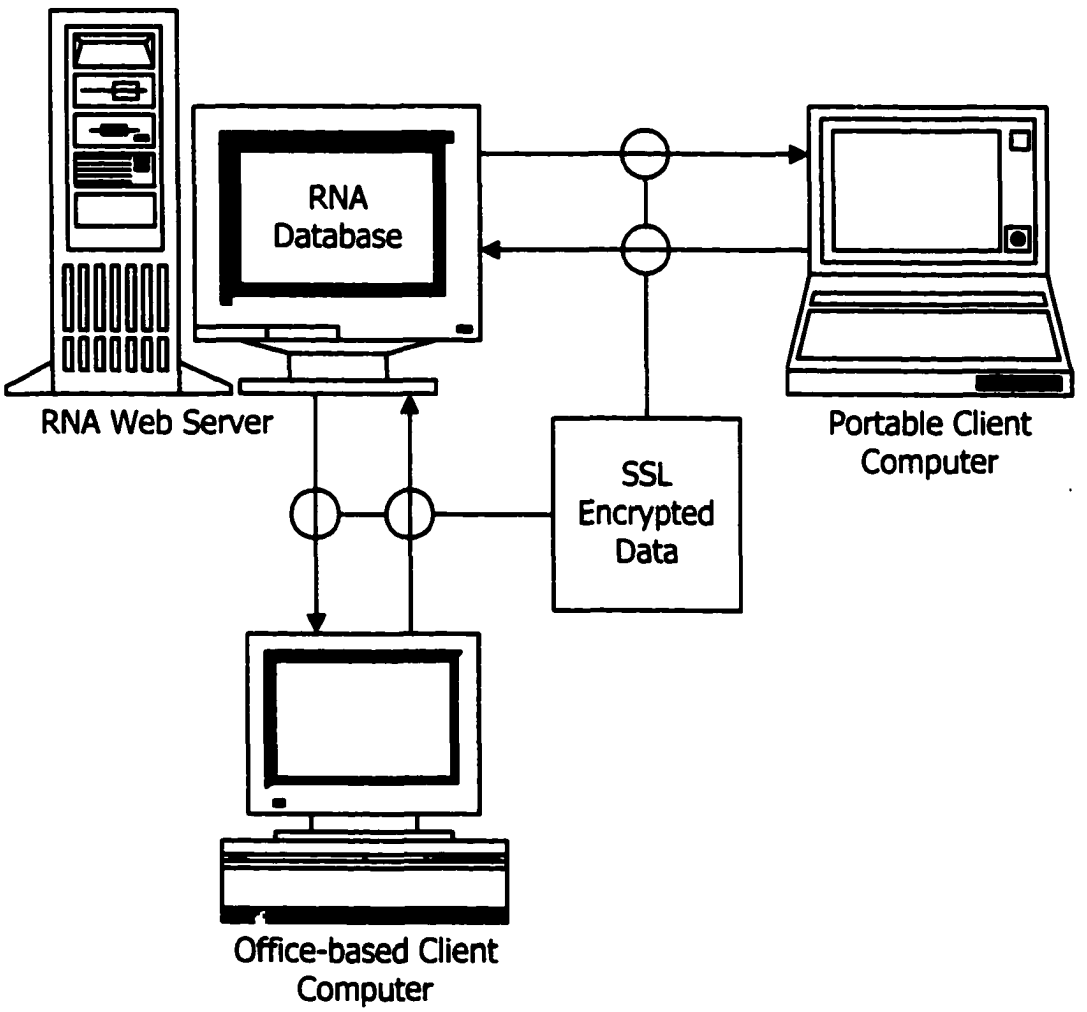


Figure 1. Remote Neuropsychological Assessment (RNA) Data Transmission Model

Figure 1 for an example of the prototype RNA system developed for the current investigation). The Microsoft Windows 95 operating system (Ver. 4.00.950) was installed on the computer, as well as a Microsoft Internet Explorer WWW browser (Ver. 4.0) modified with a Shockwave WWW browser plug-in (Macromedia Director, 1999) to "host" the RNA-CT program. A standard two-button Microsoft computer mouse acted as the primary item response modality. The RNA-CT stimuli were first created using computer graphics illustration and editing software programs, then they were imported into a multimedia creation and editing software package (Macromedia Director, ver. 7) for task construction. Careful attention was paid to approximating the stimuli characteristics of the revised HCT (Halstead & Settlage, 1943; Halstead & White, 1950; Simmel & Counts, 1957) and subsequent CT revisions (HRCT, Reitan & Wolfson, 1985; BCT, DeFilippis & Campbell, 1991). Due to the publishing of the CT stimuli, administration instructions, and rule structures in their entirety before 1978 (see Halstead & Settlage, 1943; Simmel & Counts, 1957), it can be reasonably concluded that the CT stimuli and procedures are in the public domain. The RNA-CT stimuli were reduced in size for presentation at 640 x 480 resolution on the 15" viewable area monitor size, but they did not deviate appreciably from the size characteristics of the BCT or HRCT stimuli. Auditory presentation of the RNA-CT task instructions and response feedback was channeled through two external computer speakers, which were connected to a Creative Labs PCI128 sound card installed in the computer. Sound levels

on the speakers were varied according to the listening preference of the subject, but in all cases, the sound level was loud enough to be independently perceived by the experiment administrator.

Procedure

Subject recruitment took place via undergraduate subject pool sign-up sheets posted on the first floor bulletin board of Audubon Hall at Louisiana State University (LSU). Additional recruitment efforts entailed visiting LSU undergraduate psychology classes, explaining the nature of the experiment and asking participants to volunteer in exchange for course credit. Signup sheets listed the requirement that volunteers be within the ages of 17 and 34, and remaining exclusion criteria were checked once a prospective subject arrived for the experiment. If a subject was excluded from the study, a half-hour of credit was given for the time spent with the experimenter. Coin toss randomization to groups (i.e., heads to BCT group and tails to RNA-CT group) occurred in blocks of two as the subjects arrived for the study.

Each subject was given the opportunity to ask questions about the project, then read and sign the project consent form, after which, the presentation of materials took place in the following order for both groups: 1.) neurological and psychiatric history screening (see Appendix A); 2.) Computer Anxiety Rating Scale (Rosen et al., 1992); 3.) Shipley Institute of Living Scale (Shipley, 1940); and 4.), depending upon group assignment, the BCT (DeFilippis et al., 1979; DeFilippis & McCampbell, 1991) or the RNA-CT (see Appendix B).

Subjects assigned to the BCT group were administered the BCT using the instructions provided by DeFilippis and McCampbell (1979). Subject responses were recorded on test forms provided with the BCT, and a total error score, errors per subtest, as well as total BCT completion time and time per subtest data, were derived for each subject.

The RNA-CT group was familiarized with the computer apparatus before starting the computer task. Subject hand preference was determined, and depending upon the subject's response, the computer mouse was transferred to either the right or left side of the subject to facilitate use of the dominant hand for task responding. Once a subject was comfortable with the apparatus, the experimenter accessed the remote server, called upon the WWW page in which the RNA-CT was embedded, and started the task. During task completion, the experimenter was situated behind and out-of-view of the subject, allowing the RNA-CT program to act as the primary administrator of the task. If, during completion of the RNA-CT, the subject required additional assistance from the experimenter, brief task clarification was allowed. In all cases, though, the experimenter only reiterated the instructions presented for the task, unless the nature of the subject's questions was mechanical or software error-related. Once a subject completed the RNA-CT, the experimenter saved the task data to a floppy disk and printed a hard copy of the data for archival purposes, and the WWW browser was closed and the remote server access terminated.

Upon completing the tests, both groups were debriefed about the nature of the experiment; given a copy of the consent form; and told whom to contact should they require additional information or clarification regarding their participation.

Research Model and Hypotheses

During the attempt to address the primary focus of the current investigation (i.e., does the RNA administration differ significantly from manual administration on a common neuropsychological assessment measure?), some intriguing methodological and logical difficulties came to light. As a matter of experimental logic, one cannot directly prove the null hypothesis (Popper, 1962), which is noted as an equivalence or non-significant difference between group variables (noted symbolically as, H_0: $\mu_1 = \mu_2$). Typically, null hypothesis does not present a problem, as the majority of scientific studies are more concerned with testing for the alternate hypothesis. The alternate hypothesis normally states that there is a significant relationship between study variables (symbolically noted as, H_1: $\mu_1 \neq \mu_2$), which, depending upon the study outcome, leads to a "rejection of" or "failure to reject" the null hypothesis (Christensen, 1980). In the current investigation, the veracity of the null hypothesis (i.e., equivalence or non-significant differences between groups) was the main concern, but demonstrating this directly and/or testing for a restated alternate hypothesis (i.e., testing for equivalence) was logically and methodologically problematic. The solution selected to address this logic

problem was to test for the alternate hypotheses (i.e., significant differences between groups), but with an alteration in the standard alpha level (i.e., Type I error probability). By setting the significance level for the equivalence comparisons to a p-value of .10, the analyses used were conservative in the sense that even very small differences between administration types would be detected. A generally unwanted effect of alpha level manipulation is the increased probability of committing a Type I error (i.e., stating that the groups are not equal, when they truly are equal), but this effect is offset by a reciprocal reduction in Type II error probability (i.e., stating that the groups are equal, when they are truly not equal). Given the importance of modality equivalence for the current investigation, a minimization of Type II error probability was of greater concern, as a reduced Type II error probability serves to increase the confidence that the administration types are truly equal if non-significant differences are obtained for the group comparisons.

Two levels of the equivalence were assessed between the study groups; surface characteristics, as measured by CT error scores (both total error and errors per subtest variables), and internal test characteristics, derived from the investigation of subtest error score factor structures. It was hypothesized that the RNA-CT and BCT groups would fail to show significant differences on either level, thus reasonably demonstrating the equivalence of the two administration modalities.

In addition to the primary hypotheses of CT error score and factor structure equivalence, comparisons were planned between groups on CT response timing characteristics. The use of response timing scores within the CT assessment paradigm has been sparse and not well studied (Choca & Morris, 1992; Rattan et al., 1986). The investigation of this ancillary CT measurement component was undertaken to gain a better understanding of its general utility and to establish normative response timing values for future investigation and clinical application. In keeping with the overall assumption of equivalence between the administration modality types, it was hypothesized that the groups would not differ significantly on the response timing variables.

Analytical Methods

In order to assess the level of equivalence between the test administration formats, the data from the group samples were subjected to four modes of comparison: 1) analyses of central tendency, 2) comparisons of differences in dispersion of the CT total errors variable, 3) comparisons of differences in distribution shape, and 4) comparisons of differences in internal test structure.

Analyses of central tendency were planned using t-test for independent samples statistical method on the dependent variables between groups (i.e., total error score and error scores per subtest). Error scores per subtest for subtests I – III in both groups failed to meet t-test assumptions; therefore, a non-parametric analysis procedure (Mann-Whitney U) was employed to assess for significant differences between administration formats on these select

variables. To measure group differences in variance, the Levene's test for homogeneity of variances for two independent samples was employed, and the planned comparisons for differences in distribution shape were measured using the nonparametric Kolmogorov-Smirnov test for two independent samples (D). To assess differences in shape, the RNA-CT and BCT total error and error per subtest scores were converted to z-scores allowing for a standardization of means and variance. Since the Kolmogorov-Smirnov test assesses differences in dispersion, location, and shape, standardizing the scores limited the analysis to differences in shape only (see Siegel & Castellan, 1988).

Following the evaluation of the surface characteristics of central tendency, dispersion, and shape, the study groups were then assessed for equivalence based upon the factor structure (or internal test characteristics) of each measure. The factor structure of the BCT subtest error scores was first determined with exploratory factor analysis. Once the structure of the BCT subtest error scores was ascertained, it was then possible to use the factor structure as a mode of comparison for the RNA-CT group data. Rather than relying upon simple visual inspection of the groups' factor structures to determine modality similarity or equivalence, the subtest error scores of the RNA-CT were subjected to a confirmatory factor analysis (CFA) utilizing the BCT subtest error score factor structure as the model for the underlying latent structure of the test. The goal of CFA is to discover theoretical constructs that underlie a set of observed variables by examining the covariate structures

among the observed variables, and it is utilized when evidence derived theoretically or empirically suggests that specific latent variables explain the relations between the observed variables (Ullman, 1996). CFA allows for the testing of a hypothesized statistical model against the actual set of data. The goodness-of-fit between the hypothesized model, in this case, the BCT factor structure, and the actual data set (i.e., RNA-CT error data) was evaluated using a number of inferential and descriptive statistical indices. The most common of the indices used in the current investigation include the chi-squared statistic (X^2), the goodness-of-fit index (GFI; Bentler & Bonett, 1980), the adjusted goodness-of-fit index (AGFI; Bentler & Bonett, 1980), normed fit index (NFI; Bentler & Bonett, 1980), the comparative fit index (CFI; Raykov & Marcoulides, 2000), and the Root mean squared error of approximation Index (RMSEA; Jöreskog & Sörbom, 1994).

In CFA, the X^2 statistic has been traditionally used to evaluate the fit between the hypothesized statistical model and the actual data set. The null hypothesis (i.e., no relationship between the proposed model and the data) is tested against with the CFA X^2 and the results hinge upon the extent to which the observed data covariance matrix fits the proposed model. A significant X^2, generally defined by a p-value of $\leq .05$, suggests that the proposed model does not fit the data well and model adjustment may be necessary to fit the data covariance matrix. In addition to the primary inferential X^2 index, most CFA procedures include the use of descriptive indices to offset X^2 limitations (e.g., X^2

results are negatively affected by sample size variations). Two such descriptive indices, the GFI and AGFI, provide estimates of the amount of variance and covariance that are explained by the proposed model. AGFI differs from GFI only in the inclusion of the number of model parameters when computing variance and covariance proportions. GFI and AGFI have a range of 0.00 to 1.00, and a good model fit is indicated by an GFI or AGFI that approximates 1.00. Because the X^2, GFI, and AGFI are known to be dependent on sample size, the NFI (Bentler & Bonett, 1980), which is not dependent on sample size, was computed, as well, to assist in the determination of model fit. The NFI is derived by computing the ratio of X^2 to its degrees of freedom (df) for the proposed model with a null model that specifies zero covariances among measures and sample invariant variances. Similar to the GFI and AGFI, NFI values of .90 or higher are considered to be indicators of acceptable model fit (Bentler & Bonett, 1980). The last two indices, the RMSEA and CFI, follow a similar logic to the NFI, by comparing the proposed model with a null model assuming no relationships between measures. However, they differ markedly from the other descriptive indices by assessing for noncentrality of the X^2 distribution (i.e., viewed as an index reflecting the degree to which the proposed model fails to fit the data). The RMSEA imposes weaker requirements for degree of fit between the proposed model and the data. Raykov & Marcoulides (2000) suggest that a RMSEA value of less than .05 is indicative of the proposed model being a reasonable approximation of the data.

The CFI is defined as the ratio of improvement in noncentrality (moving from the null to the proposed model) to the noncentrality of the null model (Raykov & Marcoulides, 2000). As with all of the other descriptive indices, there are no norms for how high the CFI should be in order to safely retain or reject a proposed model, but in general, a CFI in the .90s or above is usually associated with models that are plausible approximations of the data.

Once the CFA was complete, a post-hoc analysis of the RNA-CT group's subtest error scores was conducted to establish what specific internal test structure differences, if any, were detected by the CFA. This post-hoc analysis of the RNA-CT data also served as a statistical check for the CFA results.

The unique nature of the RNA-CT allowed for the collection of response time data and this information was subjected to a correlational analysis (r) to investigate the relationship between task error scores and response timing. Group comparisons using independent t-test analyses of the subtest and total CT completion times were conducted to determine if significant differences existed on this alternate CT variable.

Other methods of assessing CT administration differences were considered (e.g., reliability), but experimental design and subject pool availability were limiting factors for the current study. Prior research citing evidence of CT practice effects (DeFilippis & McCampbell, 1991; McCaffrey, Ortega, Orsillo, Nelles, & Haase, 1992; Rawlings & Crewe, 1992), made the possibility of implementing a repeated measures within-subject design improbable without

the threat of reducing CT retest error scores, especially given the fact that both the RNA-CT and BCT use similar test stimuli and scoring patterns.

Sample distribution equivalence comparison analyses were performed at the $p \leq 0.10$ significance level; whereas, the CFA and remaining supplementary analyses of the RNA-CT and BCT group data were performed at the $p \leq 0.05$ significance level. Except for the CFA analysis, which was conducted using structural equation modeling software (LISREL, ver. 8.30; Jöreskog & Sörbom, 1994), all other statistical analyses were performed using a standard statistical software package (SPSS, ver. 10.0.1). Unless otherwise stated, all hypotheses were bi-directional.

RESULTS

Overview

The current investigation concerned the assessment of equivalence between Internet-based (RNA-CT) and manual (BCT) versions of the CT (Halstead & White, 1950; Reitan, 1969; Reitan & Wolfson, 1985). Equivalence determination was performed by comparing the central tendency, dispersion, and shape of the RNA-CT and BCT error score distributions, as well as the comparison of the RNA-CT and BCT error score factor structures. The comparison of distribution characteristics was employed to assess the concurrent validity of the RNA-CT; whereas, factor structure comparisons were employed to investigate the latent constructs of the RNA-CT and BCT; thus, a test of construct validity between the two measures. The results from these two methodological approaches suggest that the RNA-CT has a relatively high level of concurrent validity and a moderate level of construct validity with the BCT.

In addition to the comparison of error scores and factor structures, the RNA-CT and BCT samples were subjected to an analysis of response timing differences. Although the CT was not originally conceived of as a timed neuropsychological assessment measure, the advent of computers and automated administration offers the prospect of an added task dimension of unknown utility (Beaumont, 1975; Rattan et al., 1986). Response timing comparisons between the RNA-CT and BCT were undertaken to explore this

novel addition to CT task paradigm and to establish a base for future investigation of CT timing factors. Comparisons of the RNA-CT and BCT total time and time per subtest responses were performed and results suggest that the RNA-CT may be a more efficient measure in terms of the overall time required for task completion. Correlational analysis of error scores and response timing variables also indicate that a positive relationship exists between the length of response timing and the number of errors committed by the subjects. This relationship was particularly strong for subjects in the RNA-CT group; whereas, little correspondence between response timing and error scores was observed in the BCT group.

Variables of Interest

For all sample distribution comparisons, a liberal alpha level of $p < .10$ was adopted to decrease the probability of a type II error (i.e., assigning equivalence between administration types when none truly exists).

To assess differences in central tendency between the RNA-CT and BCT groups, *t*-tests for independent samples were run on the majority of the dependent variables [e.g., total error and error per subtest (III-VII)] and the remaining dependent variables were assessed using a nonparametric statistical comparison. A relatively low error rate in Subtests III – V for both the RNA-CT and BCT groups resulted in a positive skew within these error score variables. Logarithmic transformations were applied to these subtest variables to meet the assumption of normality for *t*-test comparisons. The total error and error score

variables from Subtests VI and VII did not require transformation for either group. RNA-CT and BCT Subtests I and II could not be adequately transformed to allow for parametric statistical procedures, as a result Mann-Whitney U comparisons were applied to these select subtests to determine group differences in central location.

No differences between the RNA-CT and BCT groups on measures of central tendency were detected at the $p < .10$ significance level (see Table 3). Results from the t-test comparison of Subtest V errors approached the adopted .10 significance level [$t(69) = -1.37$; $p = .18$]; whereas, the remainder of the t tests on the subtest variables ranged from p-values of .432 (Subtest VI) to .99 (Subtest IV). The result of the mean total error t test comparison was non-significant, as well [$t(69) = -.411$; $p = .68$], and no significant differences in the medians were found between the RNA-CT and BCT groups on the Subtest I and II error score variables ($U = 595$; $p = .17$, and $U = 587$; $p = .47$, respectively).

To measure differences in variance, a Levene's test for homogeneity of variances for two independent samples was performed on each of the dependent variables (see Table 3). Significant differences in the variance between RNA-CT and BCT group error scores were detected for Subtest I [$t(69) = 8.496$; $p = .005$] and Subtest VI [$t(69) = 7.154$; $p = .009$] performances. Variance differences for Subtest IV and VII approached the .10

Table 3. Central Tendency, Dispersion, and Shape Differences of the Remote Neuropsychological Assessment – Category Test (RNA-CT) and Booklet Category Test (BCT)

	BCT[a]	RNA-CT[b]
Central Tendency (Ms)		
Total Errors	30.65	31.78
Errors Subtest I[c]	0.00	0.00
Errors Subtest II[c]	0.00	0.00
Errors Subtest III	7.56	6.22
Errors Subtest IV	6.82	6.84
Errors Subtest V	9.47	11.27
Errors Subtest VI	4.65	5.41
Errors Subtest VII	1.97	1.76
Dispersion (SDs)		
Total Errors	10.98	12.21
Errors Subtest I	0.00	0.23**
Errors Subtest II	0.46	0.49
Errors Subtest III	7.66	7.33
Errors Subtest IV	5.70	8.01
Errors Subtest V	5.13	5.90
Errors Subtest VI	2.75	5.07**
Errors Subtest VII	1.59	1.21
Shape (D)		
Total Errors	-----	.522
Errors Subtest I[d]	-----	-----
Errors Subtest II[d]	-----	-----
Errors Subtest III	-----	.505
Errors Subtest IV	-----	.897
Errors Subtest V	-----	.954
Errors Subtest VI	-----	.863
Errors Subtest VII	-----	1.14

Note. Central tendency measures were compared via t-test for independent samples, except for Subtests I & II errors, which were compared using Mann-Whitney U test. Dispersion characteristics were compared via Levene's test for homogeneity of variances, and the sample distribution shapes were compared via Kolmogorov-Smirnov test for independent samples (D).
[a]n = 34. [b]n = 37. [c]Median scores. [d]A general lack of variance in Subtest I & II for both groups made meaningful shape determination impossible.
**p < .01.

significance level (ps = .20 and .16, respectively), while the remaining subtest variance differences ranged from p-values of .25 (Subtest V) to .97 (Subtest III). Total error score variances were not significantly different between groups [t (69) = .039; p = .84].

Planned comparisons of the differences in distribution shape were measured using the nonparametric Kolmogorov-Smirnov (K-S) test for two independent samples (see Siegel & Castellan, 1988). To assess differences in shape, all dependent error score variables were first converted to z-scores to standardize the mean and variance. Since the K-S test measures differences in dispersion, location, and shape, standardization of the dependent variables limits the K-S test to assessing the differences in shape only. No significant differences in shape were detected for total error or error for Subtests III – VII between the RNA-CT and BCT groups (see Table 3). The determination of shape differences for Subtests I and II was impossible given the paucity of data for these variables in either group (i.e., no BCT and one RNA-CT subject made an error on Subtest I, while four BCT subjects and six RNA-CT subjects made an error on Subtest II).

The relationships between total errors and errors per subtest variables for each group were investigated using correlational analyses, and the correlations between total error and subtest error scores for the RNA-CT and BCT groups were represented in a standard matrix (see Table 4). The Spearman rank-order correlation method (Spearman rho) was selected over the Pearson r

method because of the considerable positive skew noted in the some of the errors per subtest raw scores for both the RNA-CT and BCT groups. As Table 4 illustrates, not unexpectedly, the total error score for both the RNA-CT and BCT groups correlated significantly with most of the constituent error per subtest scores. The RNA-CT group total error variable correlated most strongly with error scores from Subtest III, IV, V, VI, and VII ($ps < .01$), while the RNA-CT groups error scores for Subtest I and II clearly did not share a significant correlational relationship with the total error variable. A somewhat similar demonstration of correlational significance between the total error and error scores per subtest variables was noted in the BCT group data. The BCT group total error score variable was significantly correlated with error scores from Subtest III, IV, and VII ($ps < .01$). However, BCT group error scores for Subtest II, V, and VI failed to demonstrate a significant rank-order correlation with the total error variable. The presence of significant intercorrelation between subtest error scores was examined for both groups. Within the RNA-CT group, significant correlations were noted between Subtest III and V (r_s = -.295; p = .012), Subtest III and VI (r_s = -.276; p = .020), Subtest III and VII (r_s = .376; p = .001), and Subtest V and VI (r_s = .616; p = .000). A differing pattern of significant subtest error score correlations was found in the BCT group data. The BCT group demonstrated significant correlations between Subtest II and IV (r_s = .487; p = .003), Subtest III and VI (r_s = -.406; p = .017), Subtest III and VII (r_s = .640; p = .000), and Subtest V and VI (r_s =

Table 4. Total Error and Subtest Error Score Correlations of the Remote Neuropsychological Assessment – Category Test (RNA-CT) and Booklet Category Test (BCT)

Test Format	Error Scores							
	Total	Subtest I	Subtest II	Subtest III	Subtest IV	Subtest V	Subtest VI	Subtest VII
RNA-CT								
Total	----							
Subtest I	.118	----						
Subtest II	.061	-.086	----					
Subtest III	.452**	.090	-.050	----				
Subtest IV	.495**	.135	.151	-.032	----			
Subtest V	.394**	-.067	-.032	-.295*	-.181	----		
Subtest VI	.414**	.021	.038	-.276*	-.035	.616**	----	
Subtest VII	.535**	-.109	-.077	.376*	.231	-.006	.043	----
BCT								
Total	----							
Subtest I[a]	----	----						
Subtest II	.253	----	----					
Subtest III	.578**	----	-.081	----				
Subtest IV	.539**	----	.487**	.004	----			
Subtest V	.319	----	-.062	-.235	-.175	----		
Subtest VI	.244	----	.099	-.406*	.154	.472*	----	
Subtest VII	.666**	----	.132	.640**	.278	-.058	-.190	----

[a]Zero errors on this subtest resulted in its exclusion from the correlation analyses.
*p < .05, two-tailed. **p < .01, two-tailed.

.472; p = .005). The BCT group Subtest I error score variable was not included in the correlational analysis due to a lack of data (i.e., no errors were made by any of the BCT subjects on this task subtest). Because separate groups were used for each measure, cross-mode correlations per individual were unavailable.

In order to test the hypothesis of equivalence in factor structures between measures, the factor structure of the BCT was first determined using exploratory factor analysis, which, in turn, generated a comparison factor structure for CFA with the RNA-CT data. The total error score variable for both groups was dropped from the factor analytic investigation due to the property of singularity (i.e., total error score likely represent a linear function of the constituent subtest error scores). Examination of the BCT and RNA-CT data indicated that very few, if any, errors were made on Subtests I and II by either group. On subtest I, 100% of the RNA-CT and BCT samples made one error or less, whereas on Subtest II, 97.3% of the RNA-CT and 97.1% of the BCT samples made one error or less. These distributions do not meet the assumptions of multivariate normality desirable for factor analysis, and as a result, all subtest factor analyses were carried out with the exclusion of Subtests I and II from both RNA-CT and BCT groups. The remaining subtest error scores (i.e., Subtest III – VII) were converted to log transforms before being subjected to factor analysis in order to satisfy assumptions of normality (i.e., error scores, on the whole, were positively skewed and required

adjustment to approximate normal distribution shape). Due to the number of statistically significant correlations between subtest error score variables, an oblique (promax) rotation was performed to prevent the distortion of the resultant factor structure by forcing independence (i.e., orthogonal rotation). This decision was based, in part, upon the results of prior investigations of the CT and other neuropsychological assessment measure factor structures, which suggest that oblique rotations may add clarity to the data and permit more clinically meaningful solutions in neuropsychological measurement investigations (Donders & Strom, 1995; Livingston et al., 1996; Reynolds & Bigler, 1996).

For the BCT factor analysis, there were a total of 34 participants and 5 variables (e.g., error scores for Subtest III – VII), approximately 6.8 participants per variable. Three factors with eigenvalues of greater than 1.0 were retained for interpretation (See Table 5). The cut-off for BCT factor loadings was set relatively high at .60 to establish a homogeneous factor structure. Three factors were identified from the BCT subtest error data and are as follows: Factor 1, Simple Proportional Reasoning & Task Memory (Subtest III and VII); Factor 2, Spatial-positional Reasoning (Subtest V and VI); and Factor 3, Complex Proportional Reasoning (Subtest IV). Investigation of the factor correlation matrix suggested that the three factors extracted from the BCT data were independent components (see Table 6), and the overall BCT factor analytic results support prior research endeavors, which have suggested

Table 5. Factor Structure of the Booklet Category Test (BCT)

BCT Subtests	Factor 1	Factor 2	Factor 3
Subtest I[a]	----	----	----
Subtest II[a]	----	----	----
Subtest III	.90	----	----
Subtest IV	----	----	.96
Subtest V	----	.88	----
Subtest VI	----	.83	----
Subtest VII	.90	----	----
Eigenvalues	2.04	1.24	1.04

Note. Exploratory factor analysis (oblique rotation). Only factors with eigenvalues greater than 1.00 and loadings of .60 or greater are reported. Combined factors accounted for 86.6% of the total variance.
[a]Subtests that were excluded from factor analysis due to statistical assumption violations (e.g., lack of response data, difficulties in achieving normality assumptions).

that the CT is not a unified assessment measure, rather the test is thought to tap a number of neurocognitive skills (Allen et al., 1999; Donders & Strom, 1995; Fischer & Dean, 1990; Johnstone et al., 1997; Livingston et al., 1996).

The BCT subtests factor loadings detected in the current investigation differed somewhat from prior CT factor structure studies (Allen et al., 1999; Johnston et al., 1997; Livingston et al., 1996), but it should be noted that these investigations of internal test characteristics employed the HRCT, not the BCT, and each included error scores from Subtest I and II, which may have contributed to the disparity in CT subtest factor loadings. To date there has been no systematic investigation of the factor structure of the BCT.

Once a factor structure for the BCT was determined, it was possible to use this information to assess for a similar latent construct structure in the RNA-CT

Table 6. Interfactor Correlations of the Booklet Category Test (BCT)

Factor	1	2	3
1	----	----	----
2	-.28	----	----
3	-.00	.11	----

group via CFA; thus, directly assessing the level of construct equivalence between the two measures (Ullman, 1996). A hypothesized structural equation model (SEqM) was constructed based upon the BCT factor analytic results (see Figure 2). In Figure 2, Xs refer to indicators of the constructs; Λ (Lambda) to coefficients of the effects of the constructs on the Xs; Φ (Phi) to the variance/covariances among the latent independent variables; and δ (Theta-Delta) to error variances of the Xs. Thirteen model parameters exist in the hypothesized SEqM model; five error variances, five indicators (dependent variables), and three variance estimates among the latent (independent) variables.

The covariance matrix of the RNA-CT subtest error data was entered into a SEqM statistical computer software program (LISREL, ver. 8.30; Jöreskog & Sörbom, 1994) and syntax parameters were set to directly assess for the goodness-of-fit of the RNA-CT data to the BCT factor structure model (Raykov & Marcoulides, 2000). Based upon the overall level of equivalence between the RNA-CT and BCT on sampling distribution characteristics (see Table 3), it was hypothesized that the CFA would be non-significant, as well. In addition to the basic X^2 goodness-of-fit index, results from the SEqM analysis (see Table 7)

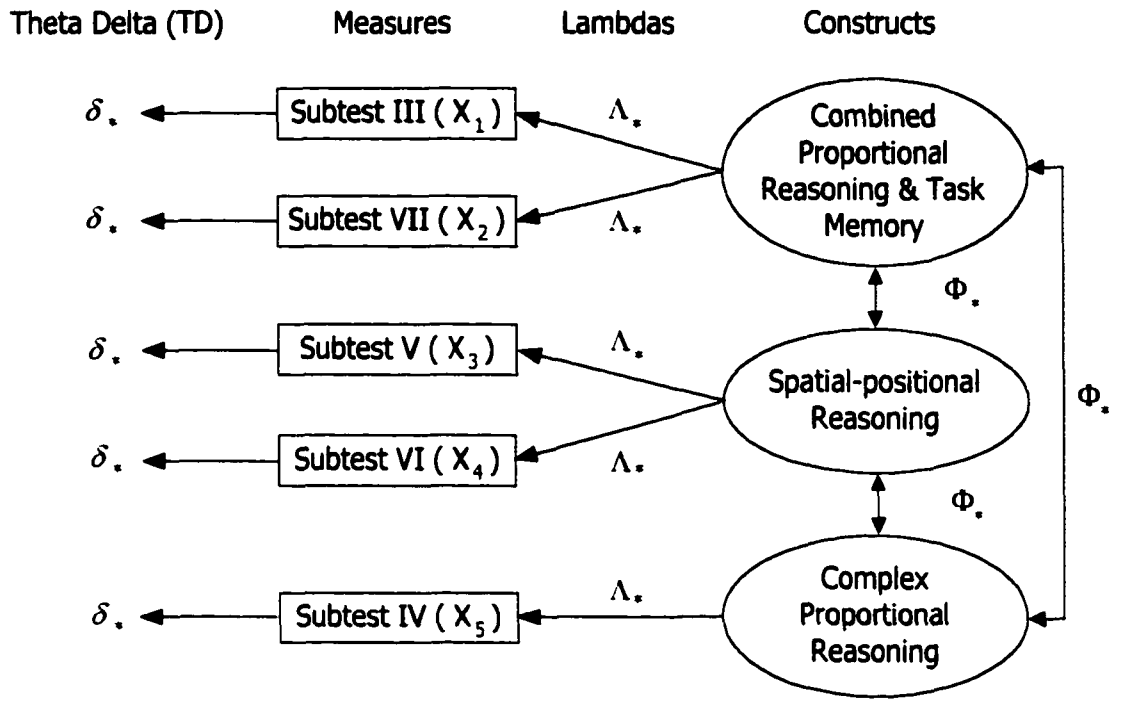


Figure 2. Hypothesized Structural Equation Model (SEqM) of the Remote Neuropsychological Assessment – Category Test (RNA-CT)

included the GFI, AGFI, CFI, NFI, and RMSEA for the BCT 3-factor model.[1] The various fit indexes for the BCT 3-factor model were mixed in their results with the bulk of the indices indicating reasonable levels of model fit for the RNA-CT group error score data. The main inferential index (X^2) value, which assessed for the null hypothesis that the proposed SEqM fits the RNA-CT covariance matrix perfectly, was statistically significant ($X^2 = 6.30$; $p = .04$). This stringent measure of model fit cannot, in isolation, be relied upon to determine model significance. The main inferential index for model fit has been long known to be adversely affected by both large and small sample sizes, as a result

[1]Readers are referred to the methodology section for more information on the various goodness-of-fit indexes employed in the current investigation.

descriptive goodness-of-fit models were also considered (Raykov & Marcoulides, 2000). GFI and AGFI results were highly suggestive for the BCT 3-Factor model (.96 and .92, respectively). It has been reported that SEqM with GFI and AGFI in the .90s or above may well represent a reasonably good approximation of the covariance data (Hu & Bentler, 1999). Similar to the GFI and AGFI, models with NFI and CFI close to a value of 1 are considered more plausible means of describing the data given the provided model than a null relationship model (i.e., a model in which no interrelationships are assumed among any of the variables). The NFI for the BCT 3-Factor SEqM was .90 and the CFI was .92, results that were at acceptable values for rejection of the null relationship model. The RMSEA approached, but was not below the suggested .05 level necessary for confident determination of model approximation (RMSEA = .07). However, investigation of the lower end of the 90% confidence interval for the RMSEA, a range of possible values for the population parameters estimated by

Table 7. Confirmatory Factor Analysis (CFA) of the Remote Neuropsychological Assessment – Category Test (RNA-CT) Structural Equivalence

Model	X^2	*df*	NFI[a]	GFI[b]	AGFI[c]	CFI[d]	RMSEA[e]
BCT 3-Factor	6.3*	12	.90	.96	.92	.92	.07

Note. Readers are referred to Bentler & Bonett (1980) and Jöreskog & Sörbom (1994) for information on the models of invariance used in the current CFA. Subtests I and II were excluded from the CFA due to concerns about statistical assumption violations.
[a]Normed Fit Index. [b]Goodness-of-fit Index. [c]Adjusted Goodness-of-fit Index. [d]Comparative Fit Index. [e]Root Mean Squared Error of Approximation Index.
*$p < .05$.

the model, suggested that the 3-Factor CFA model may be a plausible approximation for the RNA-CT error score data.

Thus, the weight of the CFA results suggests that the RNA-CT and BCT subtests measure the same latent constructs. The relatively small sample size of the RNA-CT group was a likely culprit in the conflicting results of the inferential and descriptive model fit indices; a possibility bolstered by the fact that the indices not substantiating the BCT 3-factor model are those most susceptible to sample size (i.e., X^2 and RMSEA). Bentler & Bonett (1980), as well as Raykov & Marcoulides (2000), suggest that as a general rule of thumb the covariance matrix being analyzed should derive from a sample size 10 times the number of free parameters in the proposed model. Even though the results of the CFA were equivocal and hampered by small sample size, post hoc analysis of the RNA-CT error score data substantiated the positive findings of the CFA descriptive goodness-of-fit indices. Analysis of the RNA-CT data demonstrated a factor structure similar to the BCT 3-factor solution with adequate levels of factor independence (see Tables 8 & 9). The total amount of variance given to the BCT and RNA-CT factors were comparable (e.g., BCT, 86.6%; RNA-CT, 83.6%), but differences did exist between groups in the amount of variance accounted for by the component factors. Variance differences were demonstrated in the emphasis that the Simple Proportional Reasoning & Task Memory (Subtest III and VII) and the Spatial-positional

Table 8. Factor Structure of the Remote Neuropsychological Assessment – Category Test (RNA-CT)

RNA-CT Subtests	Factor 1	Factor 2	Factor 3
Subtest I[a]	----	----	----
Subtest II[a]	----	----	----
Subtest III	----	.82	----
Subtest IV	----	----	.96
Subtest V	.88	----	----
Subtest VI	.89	----	----
Subtest VII	----	.84	----
Eigenvalues	1.96	1.21	1.08

Note. Exploratory factor analysis (oblique rotation). Only factors with eigenvalues greater than 1.00 and loadings of .60 or greater were reported. Factors accounted for 83.6% of the total variance.
[a]Subtests that were excluded from factor analysis due to statistical assumption violations (e.g., general lack of response data, normality violations).

Reasoning factors (Subtest V and VI) played in the outcome of the two measures. In the BCT group, the Simple Proportional Reasoning & Task Memory factor accounted for the most variance in error scores (40.8%); whereas, the Spatial-positional Reasoning factor took precedence in the RNA-CT group (37.5%). The differences in variance allocation between the two measures suggests that, even though the BCT and RNA-CT may share very

Table 9. Interfactor Correlations of the Remote Neuropsychological Assessment – Category Test (RNA-CT)

Factor	1	2	3
1	----	----	----
2	-.17	----	----
3	.06	-.08	----

similar factor structures, the relative emphasis that factors play in BCT and RNA-CT outcome are not equal.

Secondary Analyses

In addition to the comparison of BCT and RNA-CT error scores, analysis of response timing data was performed to investigate any possible differences between measures on this novel assessment measure component. For each subject, his/her total time for completion of task subtests was recorded minus administration time for instructions and verbal prompts; thus, the recorded total time was limited to that available to the subjects for task response and completion. In order to gain a better appreciation of time differences between subtest performances, the total response time data for both the BCT and RNA-CT groups were reduced to average time per item scores. These average response time scores were derived by taking the total time for each variable (i.e., total time, subtest I time, etc.) and dividing the total score by the number of items in the subtest(s). RNA-CT timing data was collected in milliseconds by the computer and BCT timing data was tracked by the administrator using a stopwatch and recorded in seconds. RNA-CT timing scores were converted to seconds for the purpose of group comparison and both groups' timing scores were subjected to parametric and non-parametric comparisons where appropriate (i.e., all response timing scores were subjected to independent sample t-test comparisons, except for scores from Subtests II & III which were

Table 10. Remote Neuropsychological Assessment – Category Test (RNA-CT) and Booklet Category Test (BCT) Response Time Comparisons

	BCT[a]	RNA-CT[b]
	M (SD)	M (SD)
Total Time per Subtest		
Subtest I	14.65 (4.85)	38.95 (7.82)
Subtest II	38.59 (10.75)	28.95 (7.82)
Subtest III	117.24 (53.62)	80.33 (24.91)
Subtest IV	118.00 (27.58)	88.99 (42.81)
Subtest V	119.79 (34.57)	101.91 (33.26)*
Subtest VI	94.03 (21.51)	79.90 (33.83)*
Subtest VII	47.21 (8.26)	34.82 (8.03)
Combined Subtests	549.50 (126.34)	432.55 (111.72)
Average Time per Item		
Subtest I	1.83 (.61)	3.62 (.98)
Subtest II	1.93 (.54)	1.45 (.39)
Subtest III	2.93 (1.34)	2.01 (.62)
Subtest IV	2.95 (.69)	2.22 (1.07)
Subtest V	2.99 (.86)	2.55 (.83)*
Subtest VI	2.35 (.54)	1.97 (.85)*
Subtest VII	2.36 (.41)	1.74 (.40)
Combined Subtests	2.64 (.61)	2.08 (.54)

Note. Time scores are reported in seconds and represent item response time independent of instructions or administration characteristics. Response time comparisons were conducted using t-test for independent samples, except for Subtests II & III, which were analyzed using Mann-Whitney U test.
[a]n = 34. [b]n = 37.
*All comparisons were significant at the p < .001 level, except for the total and average item response times for Subtests V & VI, which were significantly at the p < .05 level.

analyzed using the nonparametric Mann-Whitney U test due to normality violations in these two variables).

Results from the timing data comparisons indicate that the RNA-CT subjects had significantly faster response times for total test time [t (69) = 4.139; p = .000], as well as for all task subtests except for Subtest I (see Table

10). On this particular subtest, the RNA-CT subjects were actually significantly slower than the BCT subjects in task response [t (69) = -9.166; p = .000]. This finding was thought to reflect the increased time on the RNA-CT subjects' part in adjusting to the computer apparatus (e.g., mouse sensitivity and speed) and task requirements, which appears to have abated coincident with their habituation to the testing environment.

A pilot investigation of task response latency and error scores was undertaken to determine if a relationship could be ascertained between the two types of CT assessment variables for either group. Each groups' error score and response timing data were subjected to correlational analysis using the parametric Pearson's r and nonparametric Spearman's rho techniques where appropriate. Table 11 reflects the correlational relationship between error scores and response timing data for both study groups. As one investigates the correlations between response timing data and error scores, it becomes apparent that differences exist between the groups on these variables. A statistically significant positive correlation between total error and total response time was detected for the RNA-CT group (e.g., r = .54; p = .000); however, a nonsignificant inverse correlation was found for the BCT group. A lack of error score data on Subtest I precluded the correlation of subtest error score and response timing in the BCT group and no statistically significant relationships were detected between Subtest I error score and response timing for the RNA-CT group.

Table 11. Error Score and Response Timing Correlations of the Remote Neuropsychological Assessment – Category Test (RNA-CT) and Booklet Category Test (BCT)

	Response Timing Data							
Error Scores	Total	Subtest I	Subtest II	Subtest III	Subtest IV	Subtest V	Subtest VI	Subtest VII
RNA-CT								
Total	.54***							
Subtest I	----	-.05						
Subtest II	----	----	.51***					
Subtest III	----	----	----	.61***				
Subtest IV	----	----	----	----	.54***			
Subtest V	----	----	----	----	----	.63***		
Subtest VI	----	----	----	----	----	----	.74***	
Subtest VII	----	----	----	----	----	----	----	.22
BCT								
Total	-.16							
Subtest I[a]	----	----						
Subtest II	----	----	-.13					
Subtest III	----	----	----	.44**				
Subtest IV	----	----	----	----	.23			
Subtest V	----	----	----	----	----	.38*		
Subtest VI	----	----	----	----	----	----	.19	
Subtest VII	----	----	----	----	----	----	----	.12

[a]Zero errors on this subtest resulted in its exclusion from the correlation analyses.
*p < .05, two-tailed. **p < .01, two-tailed. ***p ≤ .001, two-tailed.

Similar to the disparity between total error score and total response timing found between the RNA-CT and BCT groups, Subtest II variables demonstrated the same positive correlation for the RNA-CT group (r = .51; p = .001) and a nonsignificant inverse correlation for the BCT group. Subtest III & V error and response timing variables were significantly related for both the RNA-CT and BCT groups, though the strength of the relationships were stronger in the RNA-CT group (e.g., RNA-CT ps < .001; BCT ps < .05). Error score and response timing variables for Subtests IV and VI were significantly correlated in the RNA-CT group only (e.g., Subtest IV r = .54; p = .001, & Subtest VI r = .74; p = .000). No statistically significant relationships between error score and response timing were detected in either study group for variables from Subtest VII.

DISCUSSION

Guidelines developed by American Psychological Association (APA) Committee on Professional Standards (CPS) and Committee on Psychological Tests and Assessment (CPTA) and endorsed by Division 40 (Clinical Neuropsychology) of the APA (APA, 1986; 1987) stipulate that "scores from conventional and computer administrations may be considered equivalent when (a) the rank orders of scores of individuals tested in alternate modes closely approximate each other, or (b) the means, dispersions, and shapes of the score distributions are approximately the same, or have been made approximately the same by rescaling the scores from computer mode (p. 113)." From the analysis of the current project data, it becomes clear that the BCT and RNA-CT results would satisfy the APA CPS-CPTA guidelines for measurement equivalence. Comparisons of the RNA and manual administration groups on CT total error and error per subtest variables revealed that no significant differences exist between the groups on measures of central tendency and shape, and aside from variability differences in error scores for Subtest I and VI, the differences between the dispersion characteristics between groups were negligible, as well. Bolstering the conclusion for group equivalence on the CT total error and error per subtest variables was the adopted p-value of .10 for significance, which served to increase the confidence that the groups were truly equal on the target variables (i.e., minimization of Type II error). In addition to the group comparisons based upon the surface characteristics of the CT, both

RNA-CT and BCT data were compared for similar subtest error factor structures. Confirmatory factor analysis (CFA) was used to explain the extent to which the RNA-CT and BCT measured the same underlying constructs with equivalent strength and uniqueness (Jöreskog & Sörbom, 1994). The results from the CFA indicated that the RNA-CT and BCT fit "the congeneric model" of measurement equivalence. For construct equivalence, the data must satisfy congeneric model requirements (Lord & Novick, 1968). Congeneric measures are not required to have equal error variances, so their reliabilities may not be equal. However, the factor analytic structure for congeneric tests or test batteries requires that the patterns of factor loadings be equal. Thus, the congeneric model implies that equivalent tests load on the same construct, but not with equal magnitudes. In the case of the RNA-CT and BCT subtest error score factor structures, subtest loadings [e.g., simple proportional reasoning & task memory factor (Subtests III and VII), spatial-positional reasoning factor (Subtests V and VI), and complex proportional reasoning factor (Subtest IV)] were identical between groups, but the amount of variance afford to the factors differed. In the BCT group, the simple proportional reasoning and task memory factor accounted for 41% of the total error score variance; whereas, the spatial-positional reasoning factor accounted for the majority of the error score variance (38%) in the RNA-CT group. The total error score variance accounted for by the three factors was almost identical between groups.

APA CPT-CPTA guidelines (APA, 1986; 1987) indicate that "if equivalence has been established between the conventional and computer-administered forms of the test, then the validity of the computer version can be generalized from the validity of the conventional version (p. 114)." Based upon this stipulation in the APA CPS-CPTA guidelines, it is reasonable to conclude that the RNA-CT and BCT share similar validation properties. By taking the additional step of performing a CFA and demonstrating congeneric equivalence between the RNA-CT and BCT, construct validity transfer has been reasonably determined. However, measurement equivalence and shared validity only extends to RNA-CT and BCT performances in a normal population; their equivalence in a clinical sample cannot be concluded from the current investigation. Reitan (personal communication, May 8, 2000) raises an important point addressed by Russell (1974) – most versions of the CT, including the BCT, tend to behave differently in normal and clinical populations. As a result, extended equivalence between different versions of the CT assessment paradigm can only be conferred if the measures demonstrate statistical equivalence between groups of normals and patients.

Future investigations of the RNA-CT are planned to address the issue of RNA-CT and BCT equivalence in patient populations, as well as the equivalence of the RNA-CT with other versions of the CT assessment paradigm (e.g., HRCT, SCT, Choca et al., computer version). The significant differences detected between groups on response timing variables suggests that the RNA-CT shows

promise by being a more time efficient neuropsychological assessment measure without sacrificing response accuracy, but a better understanding of what this addition to the CT assessment paradigm is actually measuring in terms of neuropsychological functioning is needed.

The present study was one of the first systematic investigations into the administration of a well-established neuropsychological assessment measure through Internet technology. The equivalence demonstrated by conventional neuropsychological assessment performance and Internet-based, RNA suggests that neuropsychological assessment procedures may be easily adapted for remote administration without incurring appreciable errors in measurement outcome. The close approximation of conventional administration procedures generated by multimedia software techniques appears to have contributed to the successful results seen in the current investigation. Future investigators of RNA should be mindful of this factor in an effort to increase the success of equivalence transfer between conventional and computerized versions of the same neuropsychological assessment measure.

RNA holds considerable promise for reaching patients who would otherwise not seek a laboratory-based neuropsychological evaluation, may be homebound, or live in inaccessible regions of the world. However, before the implementation of widespread RNA use can take place, broader measures of cognitive functioning will need to be developed, more patient samples and comparison groups will have to be examined, and normative data must be

generated specific to RNA measures. Establishing a more solid empirical basis for RNA will be vital to the longevity and effectiveness of Internet technology use in clinical neuropsychological practice in the years to come and will guarantee the attention of MCOs in search of a more efficacious means for neuropsychological assessment while keeping cost-efficiency to a maximum.

In order to flourish, the field of neuropsychology must expand and move in directions that hold promise for the expanded application of clinical and research endeavors in brain-behavior relationships. RNA is presented as one of those possible avenues, which ultimately may lead down a broken road or may hold great promise for the future of clinical neuropsychological assessment.

REFERENCES

Adams, K. M., Gilman, S., Koeppe, R., Kluin, K., Junck, L., Lohman, M., Johnson-Greene, D., Berent, S., Dede, D., & Kroll, P. (1995). Correlation of neuropsychological function with cerebral metabolic rate in subdivisions of the frontal lobes of older alcoholic patients measured with [^{18}F] fluorodeoxyglucose and positron emission tomography. Neuropsychology, 9, 275-280.

Adams, R. L., & Trenton, S. L. (1981). Development of a paper-and-pencil form of the Halstead Category Test. Journal of Consulting and Clinical Psychology, 49(2), 298-299.

Aftanas, M. S., & Royce, J. R. (1969). A factor analysis of brain damage tests administered to normal subjects with factor score comparisons across ages. Multivariate Behavioral Research, 4, 459-481.

Allen, A., & Wheeler, T. (April, 1998). Telepsychiatry: Background and activity survey. Telemedicine Today, 34-37.

Allen, D. N., Goldstein, G., & Mariano, E. (1999). Is the Halstead Category Test a multidimensional instrument? Journal of Clinical and Experimental Neuropsychology, 21(2), 237-244.

American Psychological Association (1986). Guidelines for Computer-based Tests and Interpretations. Washington, DC: Author.

American Psychological Association Committee on Professional Standards, & Committee on Psychological Tests and Assessment. (1987). Division 40: Task force report in computer-assisted neuropsychological evaluation. The Clinical Neuropsychologist, 2, 161-184.

Anderson, R. M. (1994). Practitioner's Guide to Clinical Neuropsychology. New York: Plenum Press.

Arnold, B. R., Montgomery, G. T., Castaneda, I., & Longoria, R. (1994). Acculturation and performance of Hispanics on selected Halstead-Reitan neuropsychological tests. Assessment, 1, 239-248.

Baer, L., Cukor, P., Jenike, M., Leahy, L., O'Laughlen, J., & Coyle, J. T. (1995). Pilot studies of telemedicine for patients with obsessive-compulsive disorder. American Journal of Psychiatry, 152, 1383-1385.

Ball, C., & McLaren, P. (1997). The tele-assessment of cognitive state: A review. Journal of Telemedicine and Telecare, 3(3), 126-131.

Ball, C., & Puffett, A. (1998). The assessment of cognitive function in the elderly using videoconferencing. Journal of Telemedicine and Telecare, 4, 36-38.

Ball, C., Scott, J., McLaren, P. M., & Watson, J. P. (1993). Preliminary evaluation of a low-cost videoconferencing system for remote cognitive testing of adult psychiatric patients. British Journal of Clinical Psychology, 32, 303-307.

Ball, C., Tyrrell, J., & Long, C. (1999). Scoring written material from the Mini-Mental State Examination: A comparison of face-to-face, fax, and video-linked scoring. Journal of Telemedicine and Telecare, 5, 253-256.

Barnes, G., & Lucas, G. J. (1974). Cerebral dysfunction vs. psychogenesis in the Halstead-Reitan tests. Journal of Nervous and Mental Disease, 158, 50-60.

Beaumont, J. G. (1975). The validity of the Category Test administered by on-line computer. Journal of Clinical Psychology, 31, 458-462.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of co-variance structures. Psychological Bulletin, 88, 588-606.

Berger, S. G., Chibnall, J. T., & Gfeller, J. D. (1994). The Category Test: A comparison of computerized and standard versions. Assessment, 1(3), 255-258.

Berger, S. G., Chibnall, J. T., & Gfeller, J. D. (1997). Construct validity of the computerized version of the Category Test. Journal of Clinical Psychology, 53(7), 723-726.

Bertram, K. W., Abeles, N., & Snyder, P. J. (1990). The role of learning in performance on Halstead's Category Test. The Clinical Neuropsychologist, 4(3), 244-252.

Boll, T. J. (1993). Children's Category Test. San Antonio: The Psychological Corporation.

Bornstein, R. A. (1986). Contribution of various neuropsychological measures to the detection of frontal lobe impairment. International Journal of Clinical Neuropsychology, 8(1), 18-22.

Bornstein, R. A., Baker, G. B., & Douglass, A. B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. Journal of Nervous and Mental Disease, 175, 229-232.

Boyle, G. J. (1986). Clinical neuropsychological assessment: Abbreviating the Halstead Category Test of brain dysfunction. Journal of Clinical Psychology, 42(4), 615-625.

Boyle, G. J. (1988). What does the neuropsychological category test measure? Archives of Clinical Neuropsychology, 3, 69-76.

Brandt, J., & Doyle, L. (1983). Concept attainment, tracking, and shifting in adolescent polydrug abusers. Journal of Nervous and Mental Disease, 171, 559-563.

Browndyke, J. N., Gouvier, W. D., & Waters, W. F. (1999). Predictors of ANS sympathetic activity and intensity in a large-scale MHI population: Online Neuropsychology Project II. Archives of Clinical Neuropsychology, 14(1), 26.

Browndyke, J. N., Santa Maria, M. P., Pinkston, J. B., & Gouvier, W. D. (1998). Online neuropsychology project: A survey of general head injury and prevention knowledge between professionals and non-professionals. Archives of Clinical Neuropsychology, 13(1), 133.

Burke, M. J., & Normand, J. (1987). Computerized psychological testing: Overview and critique. Professional Psychology: Research and Practice, 18(1), 42-51.

Byrd, P. B., & Ingram, C. E. (1988). A comparison study of the Intermediate Booklet Category Test with the Halstead Category Test using behaviorally disordered and normal subjects. International Journal of Clinical Neuropsychology, 10, 23-24.

Caslyn, D. A., O'Leary, M. R., & Chaney, E. F. (1980). Shortening the Category Test. Journal of Consulting and Clinical Psychology, 48(6), 788-789.

Choca, J. P. (1987). Category Test Computer Program [Computer software]. Ontario, Canada: Multi-Health Systems.

Choca, J. P., Laatsch, L., Garside, D., & Arnemann, C. (1994). Computer Category Test [Computer software]. Ontario, Canada: Multi-Health Systems.

Choca, J. P., Laatsch, L., Wetzel, L., & Agresti, A. (1997). The Halstead Category Test: A fifty year perspective. Neuropsychology Review, 7(2), 61-75.

Choca, J. P., & Morris, J. (1992). Administering the Category Test by computer: Equivalence of results. The Clinical Neuropsychologist, 6(1), 9-15.

Christensen, L. B. (1980). Experimental Methodology (2nd ed.). Boston: Allyn & Bacon.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Corrigan, J. D., Agresti, A. A., & Hinkeldey, N. S. (1987). Psychometric characteristics of the Category Test: Replication and extension. Journal of Clinical Psychology, 43(3), 368-376.

Cuevas, J. L., & Osterich, H. (1990). Cross-cultural evaluation of the booklet version of the Category Test. International Journal of Clinical Neuropsychology, 12, 187-190.

Cullum, C. M., Steinman, D. R., & Bigler, E. D. (1984). Relationship between fluid and crystallized cognitive functions using Category Test and WAIS scores. International Journal of Clinical Neuropsychology, 6, 172-174.

Dalton, J. E., Pederson, S. L., & McEntyre, W. L. (1987). A comparison of the Shipley vs. WAIS-R subtests in predicting WAIS-R full scale IQ. Journal of Clinical Psychology, 43(2), 278-279.

DeFilippis, N. A., McCampbell, E., & Rogers, P. (1979). Development of a booklet form of the Category Test: Normative and validity data. Journal of Clinical Neuropsychology, 1, 339-342.

DeFilippis, N. A., & McCampbell, E. (1991). The Booklet Category Test Manual: Research and Clinical Form. Odessa, Florida: Psychological Assessment Resources, Inc.

Desmond, D. W., Tatemich, T. K., & Hanzawa, L. (1994). The Telephone Interview for Cognitive Status (TICS): Reliability and validity in a stroke sample. International Journal of Geriatric Psychiatry, 9, 803-807.

Dodrill, C. B., & Troupin, A. S. (1975). Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. Journal of Nervous and Mental Disease, 161, 185-190.

Doehring, D. G., & Reitan, R. M. (1962). Concept attainment of human adults with lateralized cerebral lesions. Perceptual and Motor Skills, 14, 27-33.

Donders, J., & Kirsch, N. (1991). Nature and implications of selective impairment on the Booklet Category Test and Wisconsin Card Sorting Test. The Clinical Neuropsychologist, 5, 78-82.

Donders, J., & Strom, D. (1995). Factor and cluster analysis of the Intermediate Halstead Category Test. Child Neuropsychology, 1, 19-25.

Eaves, R. C., & Smith, E. (1986). The effect of the media and amount of microcomputer experience on examination scores. Journal of Experimental Education, 55, 23-26.

Eckardt, M. J., & Matarazzo, J. D. (1981). Test-retest reliability of the Halstead Impairment Index in hospitalized alcoholic and non-alcoholic males with mild to moderate neuropsychological impairment. Journal of Clinical and Consulting Psychology, 3, 256-269.

Elford, R., White, H., Bowering, R., Ghandi, A., Maddiggan, B., St. John, K., House, M., Harnett, J., West, R., & Battcock, A. (2000). A randomized, controlled trial of child psychiatric assessments conducted using videoconferencing. Journal of Telemedicine and Telecare, 6, 73-82.

Elias, M. F., Robbins, M. A., Walter, L. J., & Schultz, N. R. (1993). The influence of gender and age on Halstead-Reitan Neuropsychological Test performance. Journal of Gerontology, 48, 278-281.

Englund, C. E., Reeves, D. L., Shingledecker, T., Wilson, K. P., & Hegge, F. W. (1987). Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) I: Design and Specification of the Battery (87-10), Naval Health Research Center, San Diego, CA.

Ernst, J. (1987). Neuropsychological problem-solving skills in the elderly. Psychology and Aging, 2, 363-365.

Finlayson, M. A. J., Johnson, K. A., & Reitan, R. M. (1977). Relationship of level of education to neuropsychological measures in brain-damaged and non-brain-damaged adults. Journal of Consulting and Clinical Psychology, 45(4), 536-542.

Finlayson, M. A. J., Sullivan, J. F., & Alfano, D. P. (1986). Halstead's Category Test: Withstanding the test of time. Journal of Clinical and Experimental Neuropsychology, 8(6), 706-709.

Fischer, W. E., & Dean, R. S. (1990). Factor structure of the Halstead Category Test by age and gender. International Journal of Clinical Neuropsychology, 12(3-4), 180-183.

Fromm-Auch, D., & Yeudall, L. T. (1983). Normative data for the Halstead-Reitan Neuropsychological tests. Journal of Clinical Neuropsychology, 5, 221-238.

Golden, C. J., Osmon, D. C., Moses, J. A., Jr., & Berg, R. A. (1981). Interpretation of the Halstead-Reitan Neuropsychological Test Battery: A Casebook Approach. New York: Grune & Stratton.

Goldstein, G., & Shelly, C. H. (1972). Statistical and normative studies of the Halstead Neuropsychological Test Battery relevant to a neuropsychiatric hospital setting. Perceptual and Motor Skills, 34, 603-620.

Gregory, R. J., Paul, J. J., & Morrison, M. W. (1979). A short form of the Category Test for adults. Journal of Clinical Psychology, 35(4), 795-798.

Halstead, W. C. (1940). Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. American Journal of Psychiatry, 96, 1263-1294.

Halstead, W. C. (1947). Brain and Intelligence: A Quantitative Study of the Frontal Lobes. Chicago: University of Chicago Press.

Halstead, W. C., & Settlage, P. H. (1943). Grouping behavior of normal persons and of persons with lesions of the brain: Further analysis. Archives of Neurology and Psychiatry, 49(4), 489-506.

Halstead, W. C., & White, J. B. (1950). Manual for the Halstead Battery of Neurophysiological Tests (Mimeographed manual). Chicago: University of Chicago.

Harvey, R. J., Roques, P. K., Fox, N. C., & Rossor, M. N. (1998). CANDID – Counseling and Diagnosis in Dementia: A national telemedicine service supporting the care of younger patients with dementia. International Journal of Geriatric Psychiatry, 13, 381-388.

Heaton, R., Grant, I., & Matthews, C. (1986). Differences in neuropsychological test performance associated with age, education and sex. In I. Grant & K. M. Adams (Eds.), Neuropsychological Assessment in

Neuropsychiatric Disorders: Clinical Methods and Empirical Findings (pp. 100-120). New York: Oxford University Press.

Heaton, R., Grant, I., & Matthews, C. (1991). Comprehensive Norms for an Expanded Halstead-Reitan Neuropsychological Battery: Demographic Corrections, Research Findings, and Clinical Applications. Odessa, FL: Psychological Assessment Resources.

Hile, M. G., & Vieweg, B. W. (1996). Severely mentally ill client acceptance of automated assessments in mental health. In M. J. Miller, K. W. Hammond, & M. G. Hile (Eds.), Mental Health Computing (pp. 399-414). New York: Springer-Verlag.

Hofer, P. J. (1985). Developing standards for computerized psychological testing. Computers in Human Behavior, 1, 301-315.

Holland, T. R., & Wadsworth, H. M. (1976). Assessment of conceptual deficits in brain-damaged and schizophrenic patients. Perceptual and Motor Skills, 43, 951-957.

Holtz, J. L., Gearhart, L. P., Watson, C. G. (1996). Comparability of scores on projector- and booklet-administered forms of the Category Test in brain-impaired veterans and controls. Neuropsychology, 10(2), 194-196.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.

Huang, M. P., & Alessi, N. E. (1996). Tools for developing multimedia in psychiatry. In M. J. Miller, K. W. Hammond, & M. G. Hile (Eds.), Mental Health Computing (pp. 322-341). New York: Springer-Verlag.

Hufford, R. J., Glueckauf, R. L., & Webb, P. M. (1999). Home-based, interactive videoconferencing for adolescents with epilepsy and their families. Rehabilitation Psychology, 44(2), 1-18.

Jerome, L. W., DeLeon, P. H., James, L. C., Folen, R., Earles, J., & Gedney, J. J. (2000). The coming age of telecommunications in psychological research and practice. American Psychologist, 55(4), 407-421.

Johnstone, B., Holland, D., & Hewett, J. E. (1997). The construct validity of the Category Test: Is it a measure of reasoning or intelligence? Psychological Assessment, 9(1), 28-33.

Jöreskog, K. G., & Sörbom, D. (1994). LISREL VIII: A Guide to the Program and Applications (3rd ed.). Chicago: SPSS, Inc.

Kane, R. L., & Kay, G. C. (1992). Computerized assessment in neuropsychology: A review of tests and test batteries. Neuropsychology Review, 3(1), 1-117.

Kane, R. L., & Reeves, D. L. (1997). Computerized test batteries. In A. M. Horton, D. Wedding, & J. Webster (Eds.), The Neuropsychology Handbook: Vol. 1. Foundations and Assessment (2nd ed., pp. 423-467). New York: Springer Publishing Co.

Kelly, M. D. (1995). Neuropsychological assessment of children with hearing impairment on Trail Making, Tactual Performance, and Category Tests. Assessment, 2(4), 305-312.

Kelly, M. D., Kundert, D. K., & Dean, R. S. (1992). Factor analysis and matrix invariance of the HRNB-C Category Test. Archives of Clinical Neuropsychology, 7, 415-418.

Kent, J., & Plomin, R. (1987). Testing specific cognitive abilities by telephone and mail. Intelligence, 11, 391-400.

Kimura, S. D. (1981). A card form of the Reitan-Modified Halstead Category Test. Journal of Consulting and Clinical Psychology, 49(1), 145-146.

Kupke, T. (1983). Effects of subject sex, examiner sex, and test apparatus on Halstead Category and Tactual Performance Tests. Journal of Consulting and Clinical Psychology, 51(4), 624-626.

Laatsch, L., & Choda, J. (1991). Understanding the Halstead Category Test by using item analysis. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3(4), 701-704.

Laatsch, L., & Choca, J. (1994). Cluster-branching methodology for adaptive testing and the development of the Adaptive Category Test. Psychological Assessment, 6, 345-351.

Lange, A., van de Ven, Q. R., Schrieken, B. A. L., Bredeweg, B., & Emmelkamp, P. M. G. (2000). Internet-mediated, protocol-driven treatment of psychological dysfunction. Journal of Telemedicine and Telecare, 6, 15-21.

Lansdell, H., & Donnelly, E. F. (1977). Factor analysis of the Wechsler Adult Intelligence Scale subtests and the Halstead-Reitan Category and Tapping Tests. Journal of Consulting and Clinical Psychology, 45(3), 412-416.

Leckliter, I. N., & Matarazzo, J. D. (1989). The influence of age, education, IQ, gender, and alcohol abuse on Halstead-Reitan Neuropsychological Test Battery performance. Journal of Clinical Psychology, 45, 412-416.

Lee, J. H., Kim, J. H., Jhoo, J. H., Lee, K. U., Kim, K. W., Lee, D. Y., & Woo, J. I. (2000). A telemedicine system as a care modality for dementia patients in Korea. Alzheimer Disease and Associated Disorders, 14(2), 94-101.

Livingston, R. B., Gray, R. M., & Haak, R. A. (1996). Factor analysis of the Intermediate Category Test. Perceptual and Motor Skills, 83, 851-855.

Long, C. J., & Wagner, M. (1986). Computer applications in neuropsychology. In D. Wedding (Ed.), The Neuropsychology Handbook (pp. 548-569). New York: Springer-Verlag.

Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

MacInnes, W. D., Forch, J. R., Golden, C. J. (1981). A cross-validation of a booklet form of the Category Test. Clinical Neuropsychology, 3(1), 3-5.

Macromedia Director [Computer software, ver. 7.0]. (1999). Berkeley: Macromedia Software.

Mack, J. L., & Carlson, N. J. (1978). Conceptual deficits and aging: The Category Test. Perceptual and Motor Skills, 46, 123-128.

Matarazzo, J. D., Matarazzo, R. G., Wiens, A. N., Gallo, A. E., & Klonoff, H. (1976). Retest reliability of the Halstead Impairment Index in a normal, a schizophrenic, and two samples of organic patients. Journal of Clinical Psychology, 32, 338-349.

Matarazzo, J. D., Weins, A. N., Matarazzo, R. G., & Goldstein, S. G. (1974). Psychometric and clinical test-retest reliability of the Halstead Impairment Index in a sample of healthy, young, normal men. Journal of Nervous and Mental Disease, 158, 37-49.

Matthews, C. G., Shaw, D. J., & Klove, H. (1966). Psychological test performances in neurologic and "pseudoneurologic" subjects. Cortex, 2, 244-253.

McCaffery, R. J., Ortega, A., Orsillo, S. M., Nelles, W. B., & Haase, R. F. (1992). Practice effects in repeated neuropsychological assessments. The Clinical Neuropsychologist, 6, 32-42.

McCampbell, E., & DeFilippis, N. A. (1979). The development of a booklet form of the Category Test: A preliminary report. Clinical Neuropsychology, 1, 33-35.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114(3), 449-458.

Mercer, W. N., Harrell, E. H., Miller, D. C., Childs, H. W., & Rockers, D. M. (1997). Performance of brain-injured versus healthy adults on three versions of the category test. The Clinical Neuropsychologist, 11(2), 174-179.

Miller, D. C. (1993). Computerized Version of the Category Test for the Macintosh Personal Computer: Adult Form [Computer software]. Denton, Texas: Author.

Moehle, K. A., Fitzhugh-Bell, K. B., Engleman, E., & Hennon, D. (1988). Statistical and diagnostic adequacy of a short form of the Halstead Category Test. International Journal of Neuroscience, 42, 107-112.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., Couturier, P., & Franco, A. (1996). Psychological impact of a remote psychometric consultation with hospitalized elderly people. Journal of Telemedicine and Telecare, 3(3), 140-145.

Monteiro, I. M., Boksay, I., Auer, S. R., Torossian, C., Sinaiko, E., & Reisberg, B. (1998). Reliability of routine clinical instruments for the assessment of Alzheimer's Disease administered by telephone. Journal of Geriatric Psychiatry and Neurology, 11, 18-24.

Moses, J. A. (1985). Internal consistency of standard and short forms of three itemized Halstead-Reitan Neuropsychological Battery tests. International Journal of Clinical Neuropsychology, 7, 164-166.

Mund, A. J. (1996, December 16). Pew Study on Internet Usage. Communications Media Center at New York Law School Web Page [On-line]. Available WWW: http://cmcnyls.edu/Papers/PewStudy.htm

Nesselroade, J. R., Pedersen, N. L., McClearn, C. E., Plomin, R., & Bergeman, C. S. (1988). Factoral and criterion validities of telephone assessed cognitive ability measures. Research Into Aging, 10, 220-234.

Parsons, O. A. (1986). Overview of the Halstead-Reitan battery. In T. Incagnoli (Ed.), Clinical Application of Neuropsychological Test Batteries (pp. 155-191). New York: Plenum Press.

Parsons, O. A., Jones, B., & Vega, A. (1971). Halstead's Category Test and lateralized brain damage. Perceptual and Motor Skills, 33, 1245-1246.

Pendleton, M. G., & Heaton, R. K. (1982). A comparison of the Wisconsin Card Sorting Test and the Category Test. Journal of Clinical Psychology, 38(2), 392-396.

Perrine, K. (1993). Differential aspects of conceptual processing in the Category Test and Wisconsin Card Sorting Test. Journal of Clinical and Experimental Neuropsychology, 15, 461-473.

Pew Internet & American Life Project (2001, February 18). More Online, Doing More [On-line]. Available WWW: http://63.210.24.35/reports/pdfs/PIP_Changing_Population.pdf

Popper, K. (1962). On the Structure of Scientific Revolutions. Chicago: Chicago University Press.

Prigatano, G. P., & Parsons, O. A. (1976). Relationship of age and education to Halstead test performance in different patient populations. Journal of Consulting and Clinical Psychology, 44(4), 527-533.

Query, W. T. (1979). Category Test score as related to age in two brain-damaged populations. Journal of Clinical Psychology, 35(4), 802-804.

Rattan, G., Dean, R. S., & Fischer, W. E. (1986). Response time as a dependent measure on the Category Test of the Halstead-Reitan Neuropsychological Test Battery. Archives of Clinical Neuropsychology, 1, 175-182.

Rawlings, D. B., & Crewe, N. M. (1992). Test-retest practice effects and test score changes of the WAIS-R in recovering traumatically brain-injured survivors. The Clinical Neuropsychologist, 6, 415-430.

Raykov, T., & Marcoulides, G. A. (2000). A First Course in Structural Equation Modeling. New Jersey: Lawrence Erlbaum Associates.

Reed, H. B. C., & Reitan, R. M. (1963). Changes in psychological test performance associated with the normal aging process. Journal of Gerontology, 18, 271-274.

Reed, H. B. C., Reitan, R. M., & Klove, H. (1965). The influence of cerebral lesions on psychological test performances of older children. Journal of Consulting Psychology, 29, 247-251.

Reitan, R. M. (1955). Investigation of the validity of Halstead's measures of biological intelligence. Archives of Neurology and Psychiatry, 73, 28-35.

Reitan, R. M. (1956). Investigation of relationships between "psychometric" and "biologic" intelligence. Journal of Nervous and Mental Disease, 123, 536-541.

Reitan, R. M. (1969). Manual for Administration of Neuropsychological Tests Batteries for Adults and Children. Tucson, AZ: Neuropsychology Press.

Reitan, R. M., & Davidson, L. A. (1974). Clinical Neuropsychology: Current Status and Applications. Washington, DC: V.H. Winston & Sons.

Reitan, R. M., & Wolfson, D. (1985). The Halstead-Reitan Neuropsychological Battery: Theory and Clinical Interpretation. Tucson: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1995). Category Test and Trail Making Test as measures of frontal lobe functions. The Clinical Neuropsychologist, 9(1), 50-56.

Reynolds, C. R., & Bigler, E. D. (1996). Factor structure, factor indexes, and other useful statistics for interpretation of the Test of Memory and Learning (TOMAL). Archives of Clinical Neuropsychology, 11, 29-44.

Rosen, L. D., Sears, D. C., & Weil, M. M. (1992). Measuring Technophobia. A manual for the administration and scoring of three instruments: Computer Anxiety Rating Scale (Form C), General Attitudes

Toward Computers Scale (Form C) and Computer Thoughts Survey (Form C). California State University Dominguez Hills, Computerphobia Reduction Program.

Rothke, S. (1986). The role of set shifting cues on the Wisconsin Card Sorting Test and Halstead Category Test. International Journal of Clinical Neuropsychology, 8, 11-14.

Russell, E. W. (1974). The effect of acute lateralized brain damage on Halstead's biological intelligence factors. Journal of General Psychology, 90, 101-107.

Russell, E. W. (1992). Reliability of the Halstead Impairment Index: A simulation and reanalysis of Matarazzo et al. (1974). Neuropsychology, 6, 251-259.

Russell, E. W. (1997). Developments in the psychometric foundations of neuropsychological assessment. In G. Goldstein, T. Incagnoli (Eds.), Contemporary Approaches to Neuropsychological Assessment:: Critical Issues in Neuropsychology (pp. 15-65). New York: Plenum Press.

Russell, E. W., & Levy, M. (1987). Revision of the Halstead Category Test. Journal of Consulting and Clinical Psychology, 55(6), 898-901.

Russell, E. W., & Barron, J. (1989). A difference that is not a difference: Reply to Vanderploeg and Logan. Journal of Consulting and Clinical Psychology, 57, 317-318.

Schatz, P., & Browndyke, J.N. (1999). Trends in computer usage and means of obtaining CE credits among professional psychologists. Archives of Clinical Neuropsychology, 14(1), 159.

Schopp, L. H. (2000). Telehealth and traumatic brain injury: Creative community-based care. Telemedicine Today, 8(1), 4-7.

Schopp, L. H., Johnstone, B. R., & Merveille, O. C. (2000). Multidimensional telecare strategies for rural residents with brain injury. Journal of Telemedicine and Telecare, 6(Suppl 1), 146-149.

Shaw, D. J. (1966). The reliability and validity of the Halstead Category Test. Journal of Clinical Psychology, 22(2), 176-180.

Shipley, W. (1940). A self-administering scale for measuring intellectual impairment and deterioration. Journal of Psychology, 9, 371-377.

Shure, G. H., & Halstead, W. C. (1958). Cerebral localization of intellectual processes. Psychological Monographs: General and Applied, 72(12), 1-40.

Shute, G. E., & Huertas, V. (1990). Developmental variability in frontal lobe functioning. Developmental Neuropsychology, 6, 1-11.

Siegel, S., & Castellan, N.J. (1988). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Simmel, M. L., & Counts, S. (1957). Some stable response determinants of perception, thinking, and learning: A study based on the analysis of a single test. Genetic Psychology Monographs: Child Behavior, Animal Behavior, and Comparative Psychology, 56, 3-157.

Slay, D. K. (1984). A portable Halstead-Reitan Category Test. Journal of Clinical Psychology, 40(4), 1023-1027.

Smyth, K. A., & Harris, P. B. (1993). Using telecomputing to provide information and support to caregivers of persons with dementia. Gerontologist, 33, 123-127.

Space, L. G. (1981). The computer as psychometrician. Behavior Research Methods & Instrumentation, 13(4), 595-606.

SPSS for Windows [Computer software, ver. 10.0.1]. (1999). SPSS, Inc.

Tate, D. F., Wing, R. R., & Winett, R. A. (2001). Using Internet technology to deliver a behavioral weight loss program. Journal of the American Medical Association, 285(9), 1172-1177.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), Using Multivariate Statistics (3rd ed.; pp. 709-811). New York: HarperCollins College Publishers.

Vanderploeg, R. D., & Logan, S. G. (1989). Comparison of the Halstead Category Test and the Revised Category Test: Comment on Russell and Levy. Journal of Consulting and Clinical Psychology, 57(2), 315-316.

Vega, A., & Parsons, O. A. (1967). Cross-validation of the Halstead-Reitan tests for brain damage. Journal of Consulting and Clinical Psychology, 31, 619-625.

Watson, C. G., Thomas, R. W., Anderson, D., & Felling, J. (1968). Differentiation of organics from schizophrenics at two chronicity levels by use of the Reitan-Halstead Organic Test Battery. Journal of Consulting and Clinical Psychology, 32, 679-684.

Webster, J., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. Behavior Research Methods, Instruments, & Computers, 28(4), 567-576.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale, Revised. New York: The Psychological Corporation.

Weins, A. N., & Matarazzo, J. D. (1977). WAIS and MMPI correlates of the Halstead-Reitan Neuropsychological Battery in normal male subjects. Journal of Nervous and Mental Disease, 164, 112-121.

Wetzel, L., & Boll, T. J. (1987). Short Category Test, Booklet Format. Los Angeles: Western Psychological Services.

Williams, J.M., & Browndyke, J.N. (1996, November). Neuropsychology on the Internet. Symposium presented at the 16th Annual Meeting of the National Academy of Neuropsychology, New Orleans, LA.

Wood, W. D., & Strider, M. A. (1980). Comparison of two methods of administering the Halstead Category Test. Journal of Clinical Psychology, 36(2), 476-479.

Zachary, R. A., Crumpton, E., & Spiegel, D. E. (1985). Estimating WAIS-R IQ from the Shipley Institute of Living Scale. Journal of Clinical Psychology, 41(4), 532-540.

Zarate, C. A., Weinstock, L., Cukor, P., Morabito, C., Leahy, L., Burns, C., & Baer, L. (1997). Applicability of telemedicine for assessing patients with schizophrenia: Acceptance and reliability. Journal of Clinical Psychiatry, 58, 22-25.

APPENDIX A – NEUROLOGICAL AND PSYCHOLOGICAL SCREENING QUESTIONNAIRE

Subject Number: _____
Handedness:_____
Age:_____ _____
Education (Year in college):_____
Sex:_____
Race:_____
Color Blindness: Y N

Have you ever experienced a head injury (includes concussion) with a loss of consciousness? Y N

Number experienced:_____
For the most recent or only head injury:
When occurred:_____
Duration of unconsciousness:_____
Length of hospitalization:_____
Duration of post-traumatic amnesia:_____
Duration of retrograde amnesia:_____
Type of injury: penetrating non-penetrating

Have you ever experienced any of the following physical limitations?
Excessive clumsiness: Y N
Weakness on one side of your body: Y
N
Other:_____

Have you ever experienced a seizure of any kind? Y N

Age at first seizure:_____
Age at seizure disorder diagnosis:_____
Seizure frequency (# per week, month, or year):
Please specify:_____

Have you ever experienced a central nervous system disease? Y N

When:_____

Type of CNS disease (subtype if known):
_____ Infection
_____ Tumor
_____ Trauma

_____ Vascular
_____ Developmental
_____ Degenerative/ hereditary
_____ Toxin
_____ Metabolic
_____ Demyelinating

Have you ever had a stroke? Y N

When:_____

Type of stroke:
_____ Aneurysm
_____ TIA
_____ Infarct
_____ Hemorrhage
_____ AVM
_____ Hypertension

Have you ever received electro-convulsive shock treatment? Y N

When:_____

Have you ever used alcohol/drugs excessively? Y N

Type of drug, when used and amount used currently:
alcohol:_____
marijuana:_____
cocaine:_____
amphetamines:_____
barbiturates:_____
hallucinogens:_____

Previous Psychiatric History: Have you ever had treatment of been hospitalized for anxiety, depression, alcohol abuse or other emotional of psychological problems? Y N

If Yes - When did these problems occur?_____

Panic Disorder: Have you had times when you felt a sudden rush of intense fear or discomfort in which you felt physical symptoms such as shortness of breath, rapid heartbeat, trembling, shaking, choking sensations, nausea, dizziness, or chest pain; and / or were concerned that you would loose your mind or die?
Y N

If Yes: Have these feelings come "out of the blue" or unexpectedly? I mean that they were not associated with a particular situation and did not occur only while you were ill or under the influence of a substance such as caffeine.
Y N

How many times has this happened during the past month?_____

Post-Traumatic Stress Disorder: Have you ever experienced any extremely stressful, life threatening or traumatic event such as serious physical injury, assault, or seeing someone badly hurt or killed in the past that still troubles you? Y N

If Yes: Do you re-experience this event through "flashback" episodes or nightmares? Y N.

Do you have intrusive thoughts about this event or experience extreme anxiety in situations that remind you of this event? Y N

Social/Simple Phobia: Are you unusually afraid of objects or situations the average person is not disturbed by, such as heights, air travel, or certain animals? Y N

Obsessive-Compulsive Disorder: Are you bothered by recurrent thoughts, impulses, or images that you can't stop from coming into your mind, and which you feel are intrusive and senseless? This is not the same as worrying about things that might happen. I mean things like a parent having repeated impulses to kill a loved child, or a religious persons having recurrent blasphemous thoughts. Y N

Major Depression: Did you ever have a period of time, which lasted 2 weeks or longer, when you felt depressed, sad, hopeless, or lost interest in almost all of your usual activities? Y N

If Yes: Please explain the circumstances behind the episode:

_____

Dysthymia: In the last two years have you felt down, blue, depressed or have lost interest in things that usually give you pleasure? Y N

If Yes: Have you felt this way more days than not for most of the day during the cast two years? Y N

Mania/Hypomania: Did you ever have a period of time, at least several days long, when you felt extremely good or high-feeling very different from just being in a "good mood?" I am talking about things like feeling that you possessed special powers, having a decreased need for sleep, racing thoughts, or feeling a pressure to keep talking such that others were not able to keep up with you in conversation. Y N

Psychosis: Has there ever been a period of time when you had strange experiences such as hearing voices or seeing visions that other people could not see or hear? Y N

Has there ever been a time when people had trouble understanding you because your speech was mixed up or because you didn't make sense in the way that you were talking? Y N

Have you ever had the feeling that something odd was going on around you, that people were doing things to test you or antagonize you or hurt you so that you felt you had constantly be on guard? Y N

APPENDIX B – REMOTE NEUROPSYCHOLOGICAL ASSESSMENT – CATEGORY TEST (RNA-CT) MANUAL

Software and Apparatus Requirements

The stand alone RNA-CT (ver. 1.0) was developed for the Windows 95/98 operating system. The current version will not run on the Windows 3.11 and NT operating systems or on any version of the Macintosh operating platform. Windows 2000 operating systems will run the current version of the RNA-CT, but normative testing on this operating system has yet to be undertaken.

The Internet-based version of the RNA-CT (net ver. 1.0) will run on all operating system platforms (e.g., Windows, Macintosh, Linux, Unix), assuming that the end-user has a Shockwave-enabled World Wide Web browser. The Shockwave browser add-on is freely available from Macromedia Software (http://www.macromedia.com).

Minimum and/or Suggested Computer Requirements/Settings:

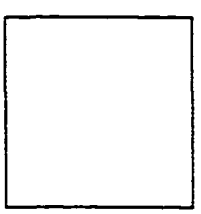
	Windows OS	Macintosh or Other OS
RNA-CT Stand Alone (ver. 1.0)	System/CPU: • 486 or Pentium processor • Windows 95 or higher Memory: • 16 MB RAM or higher recommended. Hard Drive: • 10MB of free space is recommended to store test and data materials. CD-ROM: • Necessary for test installation Audio • Sound card w/ computer speakers required Mouse/Keyboard • One or two-button mouse suggested • Keyboard response optional	Not Available Currently

	Windows OS	Macintosh or Other OS
RNA-CT Internet-based (net. ver. 1.0)	Memory: • 32 MB RAM recommended Modem: • 28.8 bps as a minimum • 56.6 bps or higher recommended. Browser: • Windows IE (ver. 3.0 or higher) or Netscape Navigator (ver. 4.0 or higher). • Macromedia Shockwave Web-browser plugin.	Memory: • 32 MB RAM recommended Modem: • 28.8 bps as a minimum • 56.6 bps or higher recommended Browser: • Windows IE (ver. 3.0 or higher) or Netscape Navigator (ver. 4.0 or higher). • Macromedia Shockwave Web-browser plugin.

Test Administration Instructions

Prior to beginning the RNA-CT, the test administrator should determine if the subject would be better suited for mouse, keyboard, or touch-screen administration. Factors such as age of the individual, prior computer mouse experience, and/or motor difficulties should guide the test administrator's decision. Keyboard, touch-screen, and mouse response modalities will work at anytime during the test, but at this time normative data is only available for mouse response times. Caution should be used in the interpretation of keyboard or touch-screen response time data until normative data has been generated for these modality types.

Once the preferred test response modality has been determined and the subject is seated in front of the computer, the administrator should begin the RNA-CT by entering the required administrator password and the unique subject identification number.



RNA-CT Introduction Screen

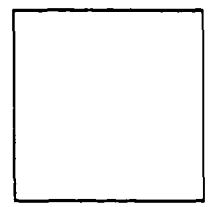
The RNA-CT test will begin. The program will display and read aloud the following text:

On the screen directly in front of you, you are going to see different geometric figures and designs. Something about the pattern on the screen will remind you of a number between 1 and 4.

Notice the buttons on the screen numbered one, two, three, and four. Using the computer mouse, you will be asked to click on the numbered buttons on the screen. You will first look at each picture on the screen, and then decide which number the picture suggests. When you figure it out, click on the numbered button on the screen that corresponds to your desired answer.

Click on the "continue" button to start the test, or click on the "repeat" button if you wish to hear the test instructions again.

The RNA-CT instructional text will clear once the subject clicks on the "Continue" arrow button at the bottom right of the screen. If further clarification of the test instructions is needed, the subject may click on the "Repeat Instructions" button at the bottom left of the screen and the text listed above will be re-read by the program.



RNA-CT Instruction Screen

Once the "Continue" button is clicked or the return key on the keyboard is pressed, the instructional text on the screen will disappear and the first test stimulus will be displayed in the area above the number buttons on the screen. The program will say:

For example, what number does this remind you of?

If the subject presses the number 1 key on the keyboard or clicks with the mouse on the number 1 button on the screen, the number 1 button on the screen will light with a green color, a bell sound will ring, and the following will be displayed and said by the program:

The bell you just heard and the green light on the button you pressed tells you that you got the right answer. Every time you have the right answer, you will hear the bell ring and a green light will appear on the button corresponding to your answer.

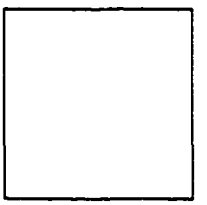
Now try clicking on another button you know is wrong for this picture.

When the subject presses or clicks on either the two, three, or four number options, a buzzer will sound and the corresponding numbered button on the screen will light with a red light. After the buzzer sounds, the program will say:

The buzzer and red light you just experienced will occur when you have the wrong answer for a picture. In this way, you will know with each picture whether you are right or wrong. However, for each picture, you get only one choice. If you make a mistake, the test will just go right on to the next picture. Let's start the test.

If the subject responds initially to the first picture with a response other than number 1, the program will say the following dialogue:

The buzzer and red light you just experienced will occur when you have the wrong answer for a picture. For this picture, you should have clicked on the number 1 button. Try clicking on the number 1 button now.



RNA-CT Tutorial Screen

After the subject presses or clicks on the 1 button, the correct response feedback sequence is initiated (i.e., bell and green button light), and the program reads the following dialogue:

The bell you just heard and the green light on the button corresponding to the number you pressed, tells you that you got the right answer. Every time you have the right answer, you will hear the bell ring and a green light will appear on the button corresponding to your answer. In this way, you will know with each picture whether you are right or wrong. However, for each picture, you get only one choice. If you make a mistake, the test will just go right on to the next picture. Let's start the test.

The second picture of subtest I will then be displayed on the screen, and the program will respond with the following dialogue:

Which number would you choose for this picture?

The program will continue to present the rest of the pictures in subtest I. After the subject has responded to the last picture in subtest I (e.g., picture #8), the following dialogue will be displayed in the test picture area of the screen and be read by the program:

That was the end of the first subtest. This test is divided into seven subtests. In each subtest, one idea or principle runs throughout the subtest. Once you have figured out what the idea or principle in the subtest, by using this idea you will get the right answer each time.

The program is going to begin the second subtest, and the idea in it may be the same as the last one or it may be different. It will be up to you to figure out.

Click on the "continue" button to start the next subtest.

When the test reaches the first slide in subtest II that contains circles (e.g., picture #9), the program will say:

You will notice that you first saw squares, then lines, and now circles. Even though the patterns change, you should continue to use the same idea to get the right answer.

The program will then continue through the pictures in subtest II. After the subject has responded to the last picture in subtest II (e.g., picture #20), the program will say:

That was the end of the second subtest and as you probably noticed, you don't necessarily have to see a number to have a number suggested to you. You saw squares, circles, and other figures. In addition, you probably noticed that in each of these subtests, there was only one idea or principle that ran throughout. Once you figured out the idea, you continued to apply it to get the right answer.

Now, the program is going to start the third subtest, and the idea in it may be the same as the last subtest or it may be different. See if you can figure out what the idea or principle is

and then use it to get the right answer. Remember, the idea remains the same throughout the subtest.

Click on the "continue" button to start the next subtest.

The program will then begin the third subtest. After the completion of the third subtest, the program will say:

That was the end of that subtest. Now, you are going to continue to the next subtest. The idea in the next subtest may be the same as the subtest you just completed or it may be different. Try to figure it out.

Click on the "continue" button to start the next subtest.

The program will continue on to the fourth subtest once the subject presses the continue key. When the test reaches the first picture in subtest IV without numerals (i.e., picture #7), the program will say:

This picture is still the same group, but now the numbers are missing. The idea or principle is still the same.

The program will continue to the end of subtest IV, after which, the following dialogue will read by the program:

That was the end of that subtest. Now, you are going to continue to the next subtest. The idea in the next subtest may be the same as the subtest you just completed or it may be different. Try to figure it out.

Click on the "continue" button to start the next subtest.

The program will then continue to subtest V, and upon completion, the program will again say the following completion dialogue:

That was the end of that subtest. Now, you are going to continue to the next subtest. The idea in the next subtest may be the same as the subtest you just completed or it may be different. Try to figure it out.

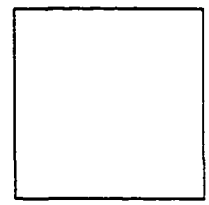
Click on the "continue" button to start the next subtest.

The program will then continue to subtest VI, and upon completion, the program will say the following dialogue:

In the last subtest, there is no one idea or principle because it is made up of items you have already seen in earlier subtests. Try to remember what the right answer was the last time you saw the picture and give that same answer again.

Click on the "continue" button to start the last subtest.

After completion of the last picture in subtest VII, the program will display an "end of test" screen indicating to the subject that the task is done.



RNA-CT End of Test Screen

See Data Retrieval and Analysis section below for instructions on how to save and analyze subjects' data.

Note – to discontinue the RNA-CT at anytime, without saving collected data:

RNA-CT Stand Alone Version (ver. 1.0) – press the Esc key
RNA-CT Internet-based Version (net. ver. 1.0) – close the WWW browser window.

Data Retrieval and Analysis

Once the test is complete, the program will prompt the administrator for the password provided at the beginning of the task. After the password is accepted, a window displaying the raw test data will appear, as well as a button allowing the administrator to save the test data as a file. Clicking on this "save data" button will generate a comma-delimited ASCII file in a computer directory called "Prefs." This "Prefs" directory will be created as a sub-directory either off the directory in which the stand-alone RNA-CT is housed or, in the case of the Internet-based RNA-CT, as a sub-directory off the web browser directory. Each test data file will be named based upon the subject ID the administrator entered at the beginning of the test and will end with the ".txt" file extension. For example, using the stand-alone version of the RNA-CT housed in a computer directory called, "c:\neuropsy\rna-ct\," a test administrator enters the subject ID "r3203." The administrator password is then provided and the test begins. Once the test is completed, the administrator enters the password

again and elects to save the test data by pressing the "save data" button. These actions create a data file called, "r3203.txt," in the computer directory "c:\neuropsy\rna-ct\Prefs\". If, in the case of the Internet-based RNA-CT version running on Microsoft Internet Explorer, the data file is saved, then the location would be "c:\Program Files\Internet Explorer\Prefs\". Date file locations for the Internet-based RNA-CT version will vary as a function of operating system, browser type, and browser installation location. Please consult your computer manual for information specific to your system.

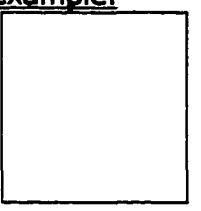
Once the data file has been saved or copied, it can easily be imported into a spreadsheet program. However, the full number of variables can only be accommodated by SPSS or SAS statistical/spreadsheet software programs. Microsoft Excel will truncate the number of variables. A Windows SPSS (ver. 10) data import template is available from the author by request (see Credit & Contact Information section), which will automate the transfer from comma-delimited data file to spreadsheet format.

The RNA-CT comma-delimited data file contains the following information (in order from the beginning to end of the file) for a total of 652 possible variables:

Subject ID	Subtest II Error Total	Subtest VI Error Total
Password	Subtest II Total Time	Subtest VI Total Time
Date of Test	Subtest II Time/Item	Subtest VI Time/Item
(month.day.year)	Subtest III Error Total	Subtest VII Error Total
Time of Test (military	Subtest III Total Time	Subtest VII Total Time
time)	Subtest III Time/Item	Subtest VII Time/Item
Total Test Error	Subtest IV Error Total	Subtest I Item 1 Subject
Total Test Time	Subtest IV Total Time	Response
Average Time Test/Item	Subtest IV Time/Item	Subtest I Item 1 (+
Subtest I Error Total	Subtest V Error Total	correct or – incorrect)
Subtest I Total Time	Subtest V Total Time	Subtest I Item 1 Time...
Subtest I Time/Item	Subtest V Time/Item	

Note – All RNA-CT response timing scores are expressed in milliseconds (ms).

Annotated RNA-CT raw data file example:



Data Retrieval and Analysis

Once the test is complete, the program will prompt the administrator for the password provided at the beginning of the task. After the password is accepted, a window displaying the raw test data will appear, as well as a button allowing the administrator to save the test data as a file. Clicking on this "save data" button will generate a comma-delimited ASCII file in a computer directory called "Prefs." This "Prefs" directory will be created as a sub-directory either off the directory in which the stand-alone RNA-CT is housed or, in the case of the Internet-based RNA-CT, as a sub-directory off the web browser directory. Each test data file will be named based upon the subject ID the administrator entered at the beginning of the test and will end with the ".txt" file extension. For example, using the stand-alone version of the RNA-CT housed in a computer directory called, "c:\neuropsy\rna-ct\," a test administrator enters the subject ID "r3203." The administrator password is then provided and the test begins. Once the test is completed, the administrator enters the password again and elects to save the test data by pressing the "save data" button. These actions create a data file called, "r3203.txt," in the computer directory "c:\neuropsy\rna-ct\Prefs\". If, in the case of the Internet-based RNA-CT version running on Microsoft Internet Explorer, the data file is saved, then the location would be "c:\Program Files\Internet Explorer\Prefs\". Date file locations for the Internet-based RNA-CT version will vary as a function of operating system, browser type, and browser installation location. Please consult your computer manual for information specific to your system.

Once the data file has been saved or copied, it can easily be imported into a spreadsheet program. However, the full number of variables can only be accommodated by SPSS or SAS statistical/spreadsheet software programs. Microsoft Excel will truncate the number of variables. A Windows SPSS (ver. 10) data import template is available from the author by request (see Credit & Contact Information section), which will automate the transfer from comma-delimited data file to spreadsheet format.

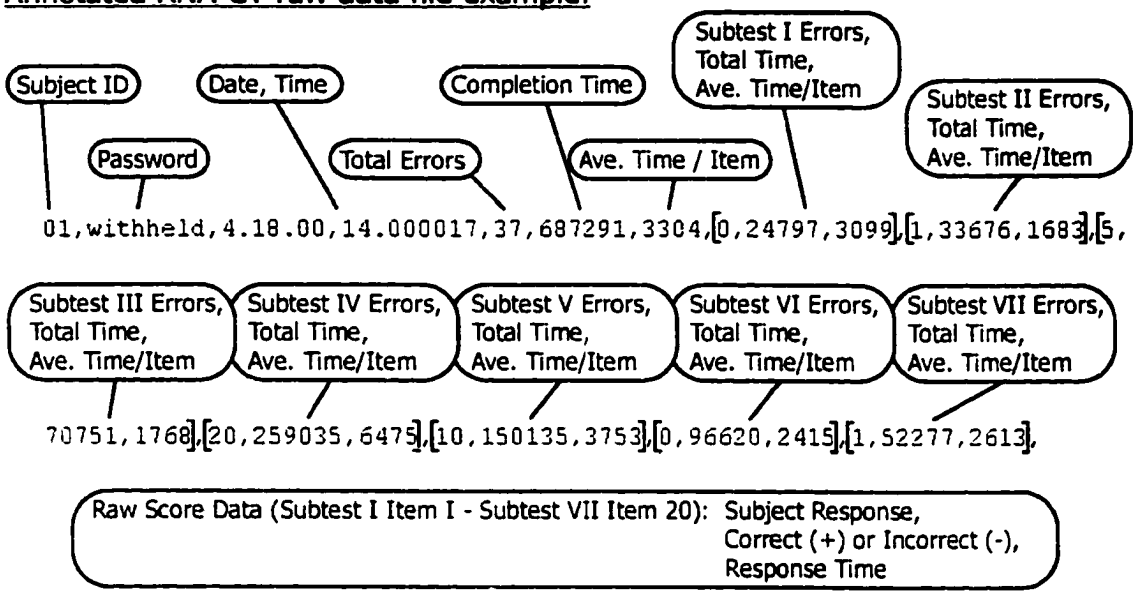
The RNA-CT comma-delimited data file contains the following information (in order from the beginning to end of the file) for a total of 652 possible variables:

Subject ID	Subtest II Error Total	Subtest VI Error Total
Password	Subtest II Total Time	Subtest VI Total Time
Date of Test	Subtest II Time/Item	Subtest VI Time/Item
(month.day.year)	Subtest III Error Total	Subtest VII Error Total
Time of Test (military	Subtest III Total Time	Subtest VII Total Time
time)	Subtest III Time/Item	Subtest VII Time/Item
Total Test Error	Subtest IV Error Total	Subtest I Item 1 Subject
Total Test Time	Subtest IV Total Time	Response

Average Time Test/Item Subtest I Error Total Subtest I Total Time Subtest I Time/Item	Subtest IV Time/Item Subtest V Error Total Subtest V Total Time Subtest V Time/Item	Subtest I Item 1 (+ correct or – incorrect) Subtest I Item 1 Time...

Note – All RNA-CT response timing scores are expressed in milliseconds (ms).

Annotated RNA-CT raw data file example:



1,+,3916,3,+,4516,1,+,2533,4,+,2033,2,+,1733,4,+,5316,1,+,1650,2,+,3100,
1,+,2166,3,+,1716,1,+,1600,4,+,1533,2,+,1500,4,+,1733,1,+,1316,2,+,1183,
3,+,1583,2,+,1883,3,+,1666,1,+,1650,4,+,1766,2,-,1733,4,+,2333,2,+,1416,
1,+,1533,4,+,1633,1,+,1933,3,+,1800,4,-,2233,3,+,2683,3,-,1583,1,-,
3066,1,-,3033,1,-,1333,1,+,5333,2,+,1400,3,+,1916,2,+,1266,3,+,1303,1,+,
1466,4,+,1633,3,+,1466,4,+,1450,2,+,966,1,+,3366,4,+,1916,1,+,1600,3,+,
1316,2,+,1650,1,+,1350,2,+,1166,4,+,1833,3,+,1866,2,+,1316,4,+,1383,3,+,
1333,1,+,2566,4,+,1216,2,+,1316,1,+,1533,3,+,1483,1,+,2000,3,+,1216,2,+,
1250,4,+,1133,3,+,1633,4,+,1550,2,+,1550,2,-,6166,3,+,4866,3,-,2750,3,-,
2650,4,-,17033,4,+,3883,1,+,14850,3,-,8200,4,-,4416,1,-,2383,4,-,7950,1,
+,12800,4,+,22250,3,+,12816,2,-,2783,4,-,13816,2,-,9833,3,-,14500,3,-,
4683,4,-,7583,4,-,13466,2,-,2933,2,+,5966,4,+,9016,3,+,5083,2,+,2866,4,
+,2466,3,+,3133,1,+,3216,2,-,4083,2,+,2383,1,+,3266,1,-,3683,1,+,1200,2,
-,1350,3,-,3216,4,+,7666,3,+,3816,4,+,1783,2,+,2233,4,-,6500,4,-,9600,3,
-,4300,1,-,2216,2,+,10383,2,-,3250,2,-,3633,2,+,2016,3,+,14166,2,+,2850,
3,+,2166,4,-,1816,4,+,2033,3,+,1716,4,+,2550,2,+,1583,1,+,2316,2,-,3950,
1,+,9650,2,-,1516,2,+,1916,1,+,1966,2,+,2833,4,+,3166,2,-,5166,2,+,4366,
4,+,1983,3,+,3316,1,+,1416,4,+,3433,2,+,4450,1,+,2766,3,+,2100,1,+,5450,
3,+,4516,2,+,3100,4,+,2000,3,+,2300,4,+,2016,2,+,3666,1,+,6450,3,+,2183,
1,+,2166,4,+,2050,2,+,2100,4,+,2116,1,+,3183,2,+,2083,3,+,1816,2,+,1416,
3,+,1700,1,+,1716,4,+,1283,3,+,4800,4,+,1600,2,+,1950,1,+,1616,4,+,1466,
1,+,2733,3,+,1283,2,+,2816,1,+,3216,2,+,1650,4,+,2116,3,+,3450,2,+,1800,
4,+,1733,3,+,2983,1,+,2116,4,+,2200,2,+,2200,1,+,3933,3,+,3716,1,+,2450,
3,+,2316,2,+,2683,4,+,1916,3,+,3100,4,+,2133,2,+,2383,1,+,2983,3,+,1783,
1,+,4883,4,+,1450,2,+,2133,4,+,1666,1,+,2966,2,+,1650,3,+,3716,2,+,1583,
3,+,2383,1,+,1916,4,+,2100,3,+,1666,4,+,4633,2,+,1833,1,+,3200,4,+,3550,
4,-,2150,3,+,4033

VITA

Jeffrey Nicholas Browndyke was born on March 31, 1969, in Memphis, Tennessee. He received the degree of bachelor of science in general psychology and biology from the University of Memphis in December 1992. Jeffrey was accepted into the psychology doctoral program at Louisiana State University in the fall of 1994. While enrolled in the medical-clinical training track, Jeffrey maintained concentrated experience in clinical neuropsychology and behavioral neurology. He was awarded the degree of Master of Arts in May of 1997. Jeffrey completed his predoctoral internship and research training at the VA Connecticut Healthcare System and Yale University School of Medicine in September of 2000. Currently, Jeffrey is a postdoctoral fellow of clinical neuropsychology within the Department of Psychiatry and Human Behavior at the Brown University School of Medicine in Providence, Rhode Island.

Jeffrey's research foci and interests include; the use of computers in clinical neuropsychology practice and assessment, neuroinformatics and psychometrics, the neuropsychological sequelae of mild traumatic brain injury and post-concussional syndrome, and the role of neuropsychology in functional neuroimaging methodologies. In addition, Jeffrey maintains an active participation in numerous professional societies and is the creator and editor-in-chief of Neuropsychology Central (www.neuropsychologycentral.com), an internationally recognized World Wide Web resource for clinical neuropsychology materials and information.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Jeffrey Nicholas Browndyke

Major Field: Psychology

Title of Dissertation: The Remote Neuropsychological Assessment-Category Test: Development and Validation of a Computerized, Internet-Based Neuropsychological Assessment Measure

Approved:

W D Gouvier
Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Sarah Pierce

Date of Examination:

2 April 2001