2012

# Examining reliability and validity of the RoadSign Perception Test

Justin Hull Ory
*Louisiana State University and Agricultural and Mechanical College*, justinhory@gmail.com

EXAMINING RELIABILITY AND VALIDITY
OF THE ROADSIGN PERCEPTION TEST

A Thesis

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
Master of Arts

In

The Department of Psychology

by
Justin Ory
B.A., Southeastern Louisiana University, 2005
M.A., Southeastern Louisiana University, 2008
May 2012

# ACKNOWLEDGMENTS

**TABLE OF CONTENTS**

# ABSTRACT

The Road Sign Perception Test (RSPT) is a newly developed assessment procedure related to driver advisement, and its psychometric properties have not yet been examined systematically. The present study examined the construct validity and the reliability of RSPT measurements. Cronbach's α and test-retest reliability statistics were completed, as were correlations reflecting on the test's construct validity. Results indicated that the RSPT possesses marginal test-retest reliability, adequate internal consistency for duration of stimulus presentation and for specific stimuli, but low for number of signs presented. RSPT performance was significantly correlated with convergent variables measuring aspects of attention, working memory, and spatial perception and orientation, all predictors of driving performance. Divergent variables measuring verbal intelligence, current mood, and handedness were not significantly correlated with RSPT performance. Since there is no normative data available for the RSPT, all of the comparisons for this study used raw accuracy scores which are naturally more variable than comparisons made between standardized scores, potentially lowering the attained reliability and setting the upper limit for observed validity. Overall, the results indicate that the RSPT will be a promising new instrument once it is completed.

# INTRODUCTION

There have been many studies examining the decline in driving ability that occurs in late adulthood (McGwin, Owsley, & Ball, 1998; McGwin, & Brown, 1999; Groeger, 2000; Evans, 2004; Hole, 2007; Eberhard, 2008; Department for Transport, 2009). This research has shown that driving ability declines naturally in late adulthood, particularly among individuals in the early stages of neurodegenerative disorders such as dementia (Reger, et al, 2004; ) or Parkinson's Disease (Uc et al, 2011). Older adults with dementia are involved in twice as many *at fault* motor vehicle accidents than a stratified random control group. Even following their collisions, many continued to drive for up to 3 or more years afterward, during which time nearly one third were involved in a second accident (Cooper, Tallman, Tuokko, & Beattie, 1993). Some studies have focused on the possible reasons for this widely observed decline (Anstey & Wood, 2011), or factors contributing to seniors' decision to "self-restrict" their own driving (Ackerman et al, 2011; Ross, Browning, Luszcz, Mitchell, & Anstey, 2011). Others have focused on evaluating driving performance with the pragmatic goal of determining at which point someone is objectively no longer safe to drive (Pachana, Fitzell, & Lie, 2004; Iverson et al, 2010).

Ultimately, the final determination whether someone is safe or unsafe to drive is made by an on-road driving test. However, on-road tests are inherently dangerous and the role of screening protocols is to ensure that prospective drivers do not have substantial impairments which would increase their risk of making a dangerous error on the road test. A number of tests and procedures have been developed to this end, ranging from the vision screening which everyone must pass to qualify for an on-road test, to neuropsychological testing, to an off-road test in a small-scale vehicle (SSV) on a closed course (Hale, Gouvier, Schweitzer, & Shipp, 1987).

A thorough review of the literature surrounding the assessment of driving advisement in late adulthood is beyond the scope of the present introduction. However, a few citations from the most recent literature are presented to characterize the current state of the field. Pachana, and colleagues (2004) examined 50 drivers aged 55-92 years who possessed current drivers licenses, each of whom was given a short battery of neuropsychological measures commonly believed to correlate with decline in driving performance. They found that longstanding personality risk factors (e.g., aggression vs. passivity, impulsivity vs. hesitancy) were poor predictors of driving ability, but certain demographics and measures of current mood and cognitive performance possessed significant predictive value.

Iverson and colleagues (2010) examined the literature for evidence-based methods of assessing driving ability in older adults. Their recommendation was that practitioners rely on the Clinical Dementia Rating Scale, caregiver rating of driving ability, history of accidents or traffic violations, self-restriction of driving under certain conditions, MMSE scores below 24, and aggressive or impulsive personality traits when making their decision whether to restrict driving. They reported that individuals' self-rating of driving ability and a lack of self-restriction are not generally useful indicators of driving ability or impairment. They reported that standard neuropsychological testing was accurate for assess driving ability in minimally vs. moderately to severely impaired adults with dementia, but that predictive ability was less accurate for borderline or mildly impaired individuals. This is consistent with the findings of Schanke and Sundet (2000), who concluded that neuropsychological testing (ie., tests of visuoconstructive ability, reaction time, visual attention, and also awareness of cognitive limitations) are good predictors of driving performance, but that on-road tests are still needed in cases of borderline findings.

**Why Current Tests Are Inadequate**

The most common methods in practice for establishing when someone is no longer safe to drive include the vision screening administered at the Office of Motor Vehicles (OMV), a physician's assessment, and neuropsychological testing to evaluate the adequacy of attention, visual perception, and cognitive processing to drive (Carr & Ott, 2010). In addition to these evaluations, there are numerous other assessment methods being developed to examine component skills involved in driving.

Although it serves as a reliable checkpoint on the way to an on-road test (Lovie-Kichin, 1988; Arditi & Cagenello, 1993), the vision screening by Snellen chart or similar test measures one ability necessary for driving, visual acuity, while other components of visual perception essential to driving ability, such as visual field, depth perception, and contract sensitivity, are not adequately assessed. Furthermore, a meta-analysis by Desapriya and colleagues (2011) did not find any studies which demonstrated the validity of vision screening as a predictor of subsequent crashes.

Physical examination by a licensed physician or other healthcare professional is useful for establishing a baseline of health and intact bodily functioning as well as identifying indicators of declining performance, but does not quantifiably measure skills necessary for driving (Carr & Ott, 2010). Neuropsychological testing was effective at identifying potentially unsafe drivers among individuals with mild or greater impairment, but predictions made based on neuropsychological assessment become less reliable among individuals with only minor impairment (Schanke & Sundet, 2000). More pragmatically, at a typical cost of $2000-$3000 per evaluation, neuropsychological assessment is expensive and time consuming. In addition, a client is only referred if there is sufficient reason to suspect impairment—such as a known

neurocognitive disorder or particularly poor performance during a physician's assessment. Therefore, although effective at separating the dangerously impaired prospective drivers from the less impaired, a comprehensive neuropsychological evaluation is prohibitive as a general screening protocol of driving ability.

There are many assessment tools currently being developed to predict driving performance. Computer-administered Sensory-Motor and Cognitive Tests (SMCTests) have predicted pass-fail rates with 70% accuracy for 200 patients with neurocognitive disorders on-road tests (Innes, Jones, Dalrymple-Alford, Severinsen, & Gray, 2009). The Screen for the Identification of Cognitively Impaired Medically At-Risk Drivers, A Modification of the DemTect (SIMARD MD) is derived from the DemTect driving assessment and was normed against an on-road driving test (Dobbs & Schopflocher, 2010).Initial validation testing indicates that SIMARD MD is a promising tool for assessing driving ability, but further evaluation has to be conducted to establish its psychometric properties.

Another driving assessment tool is the virtual reality-based driving assessment system (VR-DAS). VR-DAS utilizes a head-mounted display (HMD) to provide an immersive interactive environment to simulate the driving experience (Schultheis & Mourant, 2001). One potential benefit to using VR is that scenarios can be programmed to simulate specific driving situations which may reveal impairment (Schultheis & Rizzo, 2001). However, recent data suggests that nausea related to simulated motion is a substantial problem for subjects on the VR-DAS—even among those participants who were screened for history of motion sickness (Whipple, Mace, Manning, Ang, & Schultheis, 2011). Another major drawback is the high cost of dedicated VR equipment. Schultheis and Mourant (2001) predicted that with advancements in technology, their system may be affordable within a few years, but the last decade has not seen

4

widespread adoption of VR technology, and their only follow up to their initial report indicates a shift from driving assessment towards driving rehabilitation (Schultheis, Rebimbas, Mourant, & Millis, 2007). Virtual reality is an intriguing technology that has already been adopted by some medical professions and may one day revolutionize neuropsychology practice, but, for the present at least, VR remains only a tantalizing prospect for the future.

**Characteristics Sought in the Ideal Test**

The ideal screening instrument to detect impairment affecting driving ability should be brief, inexpensive, and easy to administer. It should not require any dedicated equipment and should be capable of being administered at the OMV counter in a way similar to how the vision screening is now administered. It should measure a number of factors observed to systematically predict driving performance, including aspects of visuoperception, attention, working memory, and executive functioning (Reger, et al, 2004; Lee, Lee, Cameron, & Li-Tsang, 2003). The Road Sign Perception Test (RSPT; Green, 2008) is a new computer-based early screening instrument for cognitive impairment. The test is designed to take 10 minutes. It will be inexpensive to use, and can be operated from any desktop or laptop computer. In addition, Dr. Green's intention when creating the RSPT was to develop a measure to detect the onset of cognitive decline before other, less sensitive diagnostic data, such as Mini-Mental State Exam (MMSE) scores, begin to reflect relevant deficits in memory or cognitive abilities (Green, 2008). Finally, the RSPT purports to measure a number of factors essential to driving ability, such as alerting attention and working memory. However, its reliability and validity have not yet been empirically established.

**Purpose and Construction of the Scale**

The RSPT tests the ability of subjects to perceive visual stimuli (up to three simultaneously presented road signs) which are presented on a computer screen either very

quickly (e.g. 25-100 msecs) for an intermediate interval of 500 msecs, or more slowly (e.g. 900-1700 msecs) (Green, 2008). In addition to target items, the test includes several distractor stimuli (i.e., pixilated target signs which should be unrecognizable) which are not accounted for in the test's scoring. The test is entirely administered by computer, with subjects using the computer mouse to select which signs were presented during each trial.

The RSPT measures iconic nonverbal memory using Canadian traffic signs. The test is designed to assess alerting attention and iconic working memory storage capacities by measuring the time in msecs a stimulus must be presented before the subject can encode it, and the number of stimuli items a client can recognize and select correctly. Both of these abilities decline naturally with age, but are especially impaired in patients with dementia (Salthouse & Babcock, 1991; Drag & Bieliauskas, 2010; see also Justino, et al, 2001; Lu, Neuse, Madigan, & Dosher, 2005; Mielke, Kessler, Fink, Herholz, & Heiss, 1995).

Although norms and standardization have not been completed yet, in developmental trials, normal, unimpaired subjects were able to get 90% of trials correct, with performance declining slightly, but not significantly, with the subject's age (Green, 2008). There were no theorized sex, occupational, or regional differences predicted by Dr. Green. At least one study of persons with right hemispheric brain damage suggests that men perform better than women on visual information processing tasks (Blanton & Gouvier, 1987), although a meta-analysis indicates that sex differences in spatial processing is dependent more on the tests examined than on broad gender differences (Voyer, Voyer, & Bryden, 1995). Also, the test uses a few uniquely Canadian road signs, a subset of which may be unfamiliar to U.S. drivers, potentially limiting their ability to effectively encode multiple signs simultaneously under time pressure, and conceptually increasing the measurement sensitivity of the test among U.S. drivers. Ultimately,

how well the RSPT detects impairment related to driving ability must be determined empirically. The present study will examine the RSPT's consistency of measurement and its ability to measure the constructs which it was intended to reflect.

**Measuring Reliability**

Reliability is the statistical yardstick of test measurement consistency (Cicchetti, 1994). According to classical test theory, every observed test score is influenced by multiple sources of variability (Anastasi, 1988). These can be categorized into the factors that contribute to test consistency, and those which contribute to test inconsistency. The true score (i.e., a representation of all the factors which contribute to consistency in the measurement) on any test is the difference between the attained score on the test, and the error of measurement (i.e., a representation of discrepancies between obtained scores and the corresponding true scores). Measurement errors across multiple individuals and test administrations are assumed to be essentially random. If errors are essentially random variables, then their effects will statistically cancel each other out, and as such, they should not correlate with true scores or errors on other tests. Measurements of reliability attempt to provide an estimate of the relative influence of error and true scores, as well as test variance across different test administrations. (Murphy & Davidshofer, 2005). Small variance in observed scores does not interfere with a test's ability to measure a subject's true test score; however, greater levels of unsystematic measurement error can spoil a test, making its results uninterpretable or relatively meaningless.

There are several ways to measure a test's reliability. For the present study, the author will assess the RSPT's test-retest reliability, and its internal consistency using Cronbach's α. Test-retest reliability is an estimate of reliability that involves comparing two administrations of the same test for the same subject (Mitchell, 1979). Because the same test is administered twice

to the same participant, the true score should be relatively constant.  Therefore, any variance

must be due to measurement error, and any systematic variance must be due to systematic error.

Test-retest is the simplest method for measuring reliability, but is vulnerable to reactivity (e.g.,

reactions to taking the test such as changes to test-taking strategy, fatigue, etc.) and memory

effects (e.g., if the subjects remember their responses from the previous administration).  Thus a

second measure, Cronbach's α, is also used to represent the test's internal consistency and

unidimensionality.  This statistic represents the overall degree to which each individual item

relates to the test as a whole, and can be computed based on the results of a single administration.

Test-retest reliability will be obtained by allowing participants to take and retake the

RSPT, and comparing the retest performance to the original test performance.  Each participant's

scores on the first administration should be correlated with their scores on the second

administration.  Both scores came from the same individual, so barring practice effects and

strategy changes they should be identical.  And, because influences on measurement error are

assumed to be mostly random between first and second administrations (i.e., with a null mean

influence), any remaining variance should be the result of systematic error.  Therefore, the

stronger the correlation between first test and second test the higher the test-retest reliability.

However, because the two scores are necessarily confounded by time, test-retest reliability is

subject to fatigue and practice effects.  Therefore, an additional method of measuring reliability

independent of time should be used to compliment test-retest reliability.

Cronbach's α will be measured for the first administration of the RSPT only.  Because it

is only necessary to administer the test once to allow computation of α, this method avoids the

theoretical problems associated with test-retest reliability as the subjects have no prior

experience to react to or remember (Murphy & Davidshofer, 2005).  However, α is also

influenced by factors other than the test's operating characteristics, and these factors warrant caution when interpreting the α statistic: a) α will be lower when the test is administered to a broadly heterogeneous population, and will be higher when evaluated using a more homogenous sample; b) α is affected by the length of the scale, so that longer tests may achieve satisfactory internal consistency despite low inter-item correlations; c) where α is extremely high, it may indicate redundancy among test items, suggesting the need to shorten the measure for the sake of parsimony and eloquence, as well as to reduce the opportunities for error variance influence to act (Streiner, 2003).

**Measuring Validity**

Validity is the statistical measure of a test's ability to measure what it is intended to measure (Murphy & Davidshofer, 2005). As with reliability, there are many ways to measure validity. For the present study, the author will evaluate construct validity, which is a measure of how well observed relationships between test constructs match those predicted by some theory (Cronbach & Meehl, 1955). In this case, the authors predicted that the RSPT measures alerting attention and working memory, so in theory it should positively correlate with measures of attention and working memory, as well as visuospatial ability as the test makes heavy use of largely nonverbal, visuospatial processing, requiring subjects to correctly perceive and recognize symbols from a list (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). However, the test should be relatively less correlated with measures of verbal intelligence or handedness. If there is a correlation between mood and attention, the test may also be sensitive to the subject's affect. If so, it represents an undesirable extraneous influence on the test rather than a theoretical stumbling block, as the effect of affect could be measured and controlled for by statistical covariance or the development of affect-corrected norms tables or an adjustment factor.

Construct validity is a measure of how well observed relationships between test constructs match those predicted by theory. The two types of relationships needed to demonstrate a test's construct validity are convergent relationships (i.e., relationships predicted to correlate rather strongly), and discriminant relationships (i.e., relationships predicted not to correlate as strongly, if at all). Construct Validity can be established by comparing expected patterns of predictable relationships using a nomological network (i.e., a map of associated factors and their predicted relationship—either convergent or discriminant, to the factors under investigation—See Figure 1).

The three factors predicted to be measured by the RSPT are alerting attention, working memory, and visuospatial perception. As such, other measures of these abilities should be positively correlated with the RSPT (convergent validity). To establish a convergent factor for attention, 2 subtests of the Test of Everyday Attention (TEA): Map Search (sustained, selective attention), and Visual Elevator (Accuracy Score: sustained attention, attentional switching; Timing Score: attentional switching, processing speed) will be used (Robertson, Ward, Ridgeway, Nimmo-Smith, 1994). To establish the convergent factor of working memory, 2 subtests of the Wechsler Memory Scale, 3$^{rd}$ edition, will be administered: Digit Span and Spatial Span. These will be used to examine 3 aspects of working memory: verbal working memory (i.e., digit span), visual working memory (i.e., spatial span), and a general composite reflecting working memory manipulation (i.e., "∑backwards" or the sum of digit span backwards and spatial span backwards). To establish a convergent factor for visuospatial processing, three brief neuropsychological assessments will be used: the Benton Judgment of Line Orientation Test (JLOT), the Benton Facial Recognition Test (FRT), and the Hooper Visual Organization Test (VOT) (Benton, 1983; Hooper, 1958).
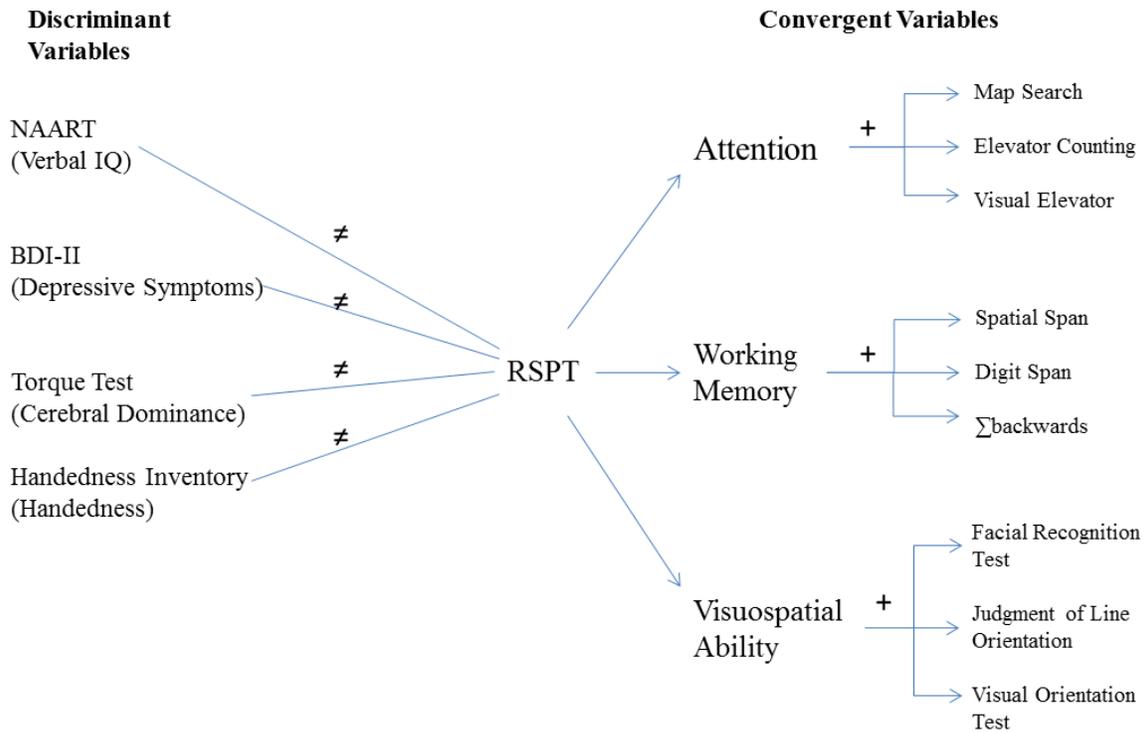
Figure 1. Nomological network of predicted relationships between the Road Sign Perception Test and other measures of attention, memory, visuospatial ability, and discriminant variables.

Figure 1 note: Discriminant variables (to left) are predicted not to correlate, or to correlate to a lesser degree than convergent variables (to right). Subdivisions of each convergent variable measure distinct aspects of the variable. Any of which may or may not correlate more closely with the RSPT that the overall variable.

Discriminant validity can be established by comparing the RSPT to theoretically unrelated factors such as handedness, verbal language ability, and (hopefully) mood. Handedness will be measured using a handedness questionnaire based on the Edinburgh Inventory (Oldfield, 1971) and the Torque Test (Blau, 1974; Blau, 1991). Because vocabulary and reading ability predict performance on many dementia screening tests (Manly, Jacobs, Touradji, Small, & Stern, 2002), we will also administer the North American Adult Reading Test (NAART), which should be unrelated to the more visually oriented RSPT. Finally, depression is the most common mental disorder among people over 65, and several of the symptoms (e.g., psychomotor agitation, diminished ability to think or concentrate, etc.) overlap with deficits of dementia (Epstein, 1976), which could potentially confound the results of the RSPT for examinees with depressive symptoms. Therefore, the Beck Depression Inventory, 2$^{nd}$ edition, will be given to each participant to estimate the effects of mood on the RSPT.

**Hypotheses**

Performance on the RSPT at test and retest should correlate very highly. A test-retest reliability coefficient (e.g., Pearson's r) of .7 is considered adequate and a test-retest coefficient greater than .9 is considered very good. For internal consistency, three measurements of Cronbach's α will be taken, examining performance at various time intervals ($\alpha_{interval}$), between differing number of stimuli ($\alpha_{number}$), and between the various road signs used in the test ($\alpha_{sign}$).

Because excellent internal consistency may be evidence of redundancy in a test, an excessively high α is also undesirable, so an α between .7 and .9 is preferred. It is hypothesized that the RSPT will demonstrate adequate reliability, both by test-retest correlation and by all three of our measurements of Cronbach's α.

$H_{\text{test-retest}}$: $r_{tt} > .7$

$H_{\alpha 1}$: $.7 < \alpha_{ms} < .9$

$H_{\alpha 2}$: $.7 < \alpha_{123} < .9$

$H_{\alpha 3}$: $.7 < \alpha_{sign} < .9$

The test items of the RSPT are hypothesized to vary in difficulty across two independent variables (IV): the number of stimuli to be encoded, and the duration of the stimulus presentation time. These two IVs are hypothesized to reflect two factors respectively: working memory and alerting attention, along with a third factor—visuoperception, which in principle is held relatively constant across each item due to the visual modality of the test. Therefore, the degree to which other tests of working memory, attention, and visuoperceptual ability correlate with participant's performance on the RSPT demonstrates the instrument's construct validity by showing positive correlations to reflect convergent validity. In addition, the degree to which theoretically unrelated factors, such as verbal intelligence, current mood, and handedness, are significantly less correlated with performance on the RSPT serves as another gauge of construct validity by demonstrating discriminant validity. It is speculated that mood might correlate with performance on the RSPT, but it is hoped that it will not.

It is hypothesized that performance on the RSPT will correlate significantly with measures of theoretically related constructs such as attention, working memory, and visuospatial ability, Performance on the RSPT will not correlate significantly with measures of theoretically unrelated constructs, including verbal IQ, mood, handedness.

$H_{V1}$: $r_{\text{RSPT1-ATTN}} > r_{\text{RSPT1-disc}}$

$H_{V2}$: $r_{\text{RSPT1-WM}} > r_{\text{RSPT1-disc}}$

$H_{V3}$: $r_{\text{RSPT1-VIS}} > r_{\text{RSPT1-disc}}$

## METHOD

**Participants**

Participants were recruited from undergraduate psychology classes, and received extra credit for their participation. Two participants were excluded from out analyses because they did not complete the assessment battery. In total, 98 participants (26 males, 72 females) completed the assessment battery. All were between 18-42 years of age ($\mu = 20$, SD = 3.35). In addition, 3 participants did not complete the Torque Test, and so were not included in the analyses for that test.

A further 6 participants were excluded from final analyses because embedded indices of suboptimal effort on the RSPT indicated that their performance on one or more administrations of the RSPT was greater than 2 standard deviations below the mean performance for healthy controls from preliminary normative data, and up to a standard deviation below the mean for confirmed malingerers. The influence of these excluded participants, who were all statistical outliers (Figure 2), was included in Table 1, in the interest of transparency of data. Leaving a final sample of 92 participants (Sex: 24 males, 68 females; Age: $\mu = 20$, SD = 3.39) in the analyses.

**Materials**

Participants completed four test batteries—three measuring theoretically distinct constructs (i.e., working memory, attention, and visuospatial ability), and a fourth, discriminant, battery consisting of measures of theoretically unrelated constructs. The working memory battery consisted of the Digit Span and Spatial Span subtests of the Wechsler Memory Scale, 3rd Ed (WMS-III; Wechsler, 1997). The attention battery consisted of the Map Search and Visual Elevator subtests from the Test of Everyday Attention (TEA; Robertson, et al, 1994). The
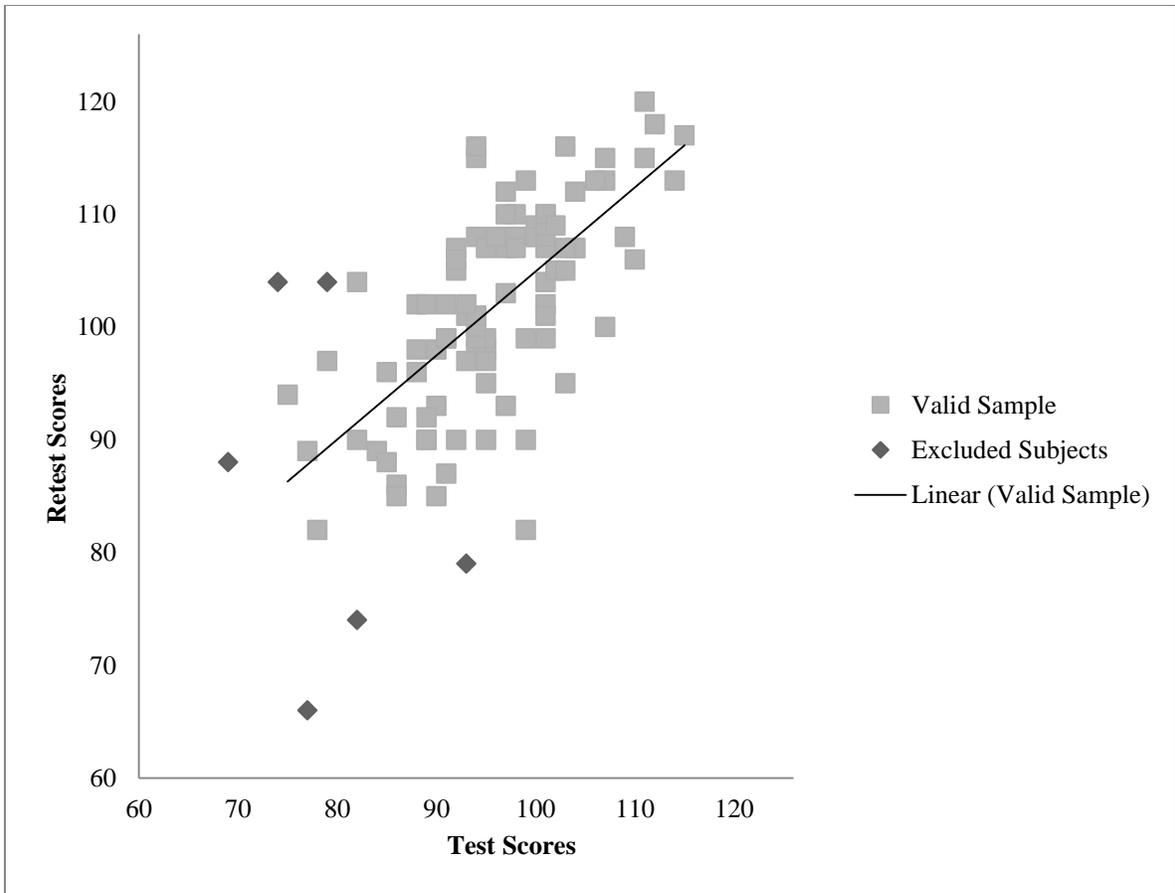
Figure 2. Scatterplot of raw score totals at test and retest for each participant (n = 98).

Table 1. Results of a bivariate correlation between RSPT and measures of attention, working memory, visuoperception, language, mood, and handedness for the full sample (n = 98), and excluding participants whose performance indicated suboptimal effort (n = 92).

| | Pearson's r with $RSPT_1$ | | | |
|---|---|---|---|---|
| Subtests by primary factor | n = 98 | n = 92 | Change | Effect Size[a] |
| $RSPT_2$ | .658** | .685** | + | Large |
| Map Search 2 minutes | .167 | .218* | + | Small |
| Visual Elevator-accuracy | .228* | .202 | - | Small |
| Visual Elevator-timing | .071 | .079 | = | Negligible |
| Spatial Span | .294** | .254* | - | Small |
| Digit Span | .251* | .262* | + | Small |
| ∑ spatial + digits backwards | .403** | .372** | - | Medium |
| Facial Recognition Test | .114 | .134 | + | Small |
| Judgment of Line Orientation Test | .206* | .276** | + | Small |
| Visual Orientation Test | .174 | .214* | + | Small |
| NAART | .193 | .194 | = | Small |
| Beck Depression Inventory-II | .048 | .034 | - | Negligible |
| Torque Test | .171 | .117 | - | Small |
| Handedness Inventory | .184 | .178 | = | Small |
| Age | -.110 | -.133 | + | Small |
| Sex | .263** | .366** | + | Medium |

Significance at .05 level = * ;  significance at .01 level = **
[a] Cohen, J (1992). A power primer. *Psychological Bulletin*, *112* (1), 155–159

visuospatial battery consisted of the Hooper Visual Organization Test (VOT; Hooper, 1958), the Benton Facial Recognition Test (FRT) and the Benton Judgment of Line Orientation Test (JLOT; Benton, 1983).

The discriminant variable battery included the North American Adult Reading Test (NAART; Strauss, et al, 2006), to estimate verbal ability; the Beck Depression Inventory, 2nd edition (BDI-II; Strauss, et al, 2006), to estimate depression or mood disturbance; and a handedness battery consisting of a modified handedness questionnaire (Cohen, 2008) and Torque Test (Blau, 1991). Participants also completed the Road Sign Perception Test (RSPT; Green, 2008) at the beginning and end of the testing session.

**Research Design and Procedure**

Each participant completed the RSPT first, followed by the four test batteries (i.e., working memory, attention, visuospatial, and discriminant), and finished with a second administration of the RSPT. The order of presentation of the four test batteries was

systematically varied between subjects using a Latin square to control for potential order of test presentation effects. For each subtest, standard manualized test procedures were followed.

## RESULTS

### Reliability Analyses

The results of a bivariate correlation of test and retest performance on the RSPT indicated marginal test-retest reliability ($r_{tt}$ = .685) for the sample. Generally, a .70 test-retest correlation is considered the lower limit of adequate (Strauss, et al, 2006). However, since the present analysis used raw accuracy scores, rather than norm-referenced standard scores which control for random error, it is likely this is an underestimation of the instrument's ultimate test-retest reliability. For comparison, the test-retest reliability of the digit span subtest of the WAIS-III, which has high temporal stability for norm-referenced scaled scores, has a .70 test-retest reliability for raw score performance in a sample of high school student athletes (Barr, 2003).

Cronbach's α indicated adequate internal consistency when examining accuracy at varying stimulus presentation time intervals ($\alpha_{ms}$ = .780), indicating that variance at each interval was small in relation to overall variance across the various time intervals. Measurement of internal consistency for accuracy of recognition when number of signs varied was low ($\alpha_{123}$= .473), indicating that variance within one or more conditions was a larger proportion of overall variance than was hoped. This may have been due in part to the influence of the distractors, which are not presently factored into the performance statistics, but would make a 1 stimulus trial appear to be a two stimulus trial, a two stimuli trial appear to be a three stimuli trial, and a three stimuli trial appear to be a four stimuli trial, and could increase the variance within each category, while reducing the variance between each category.

Finally, internal consistency was examined for recognition of the types of signs presented (e.g., Airport, Stop, Yield, etc.). Results of Cronbach's α for type of sign was adequate, ($\alpha_{sign}$ = .724), indicating that variance for each sign was small in relation to overall variance between

signs.  Overall, these findings indicate that the RSPT is capable of achieving an acceptable

degree of reliability on which to confidently base neuropsychological decisions.

**Validity Analyses**

The construct validity of the RSPT was evaluated using a bivariate correlation matrix

(Pedhazur, 1997).  The RSPT's two IVs: the number of stimuli to be encoded and the duration of

the stimulus presentation time, as well as its predominantly visual modality, are hypothesized to

reflect three factors: working memory, alerting attention, and visuoperception.  In addition,

theoretically unrelated factors such as verbal intelligence, current mood, and handedness were

assessed for each subject, and compared to RSPT performance to demonstrate discriminant

validity.  Hypotheses are depicted most clearly in the nomological network in Figure 1.

Refer to Table 1 for a concise breakdown of correlations and effect sizes.  Overall,

correlations were generally smaller than were expected for convergent variables.  Discriminant

variables correlated to a lesser extent than the convergent variables, but more so than was

expected.  The smaller than expected correlations for convergent variables is likely an artifact of

the tests chosen to represent convergent variables.  Rather than selecting tests based on the

hypothetical attentional, memory, and visuoperceptual processes believed to be measured by the

RSPT, the author selected attention tests which sampled an array of subprocesses, with the intent

of identifying which correlated most closely with RSPT performance.  As an unintended

consequence, none of the measures selected assessed for alerting attention specifically.  The

divergent measures, on the other hand, were selected in anticipation of theoretically-relevant

confounds to the RSPT and, perhaps ironically, did a better job of matching their intended

constructs than did the convergent measures.  In spite of these design flaws, correlations were

generally significant between convergent variables and the RSPT, and were not significant between discriminant variables and the RSPT.

Within the Attention battery, total number of signs correctly located within two minutes on the Map Search task was significantly correlated with raw accuracy on the RSPT (Map Search-2 min: r = .218, $p < .05$).  Accuracy on the Visual Elevator task and a measure of seconds per attention switch on the Visual Elevator task was not significantly correlated with total raw score on the RSPT (Visual Elevator-Accuracy: r = .202, $p > .05$; Visual Elevator-Timing: r = .079, $p > .05$).  Since none of the subtests selected measured alerting attention, the small but significant correlations between the RSPT and measures of other attention processes cannot fully support the hypothesis that the RSPT is measuring attention.

Within the Working Memory battery, all three measures were significantly correlated with raw accuracy on the RSPT (Spatial Span: r = .248, $p < .05$; Digit Span: r = .262, $p < .05$; $\sum$backwards: r = .372, $p < .01$).  This supports the hypothesis that the RSPT is measuring some components of working memory.

Within the Visuospatial battery, the JLOT and the VOT were significantly correlated with raw accuracy performance on the RSPT (JLOT: r = .276, $p < .01$; VOT: r = .214, $p < .05$).  This supports the hypothesis that the RSPT measures visuospatial ability.  There was also a small but insignificant correlation between the FRT and the RSPT (FRT: r = .134, $p > .05$), indicating that subjects were using line orientation and identification of distinctive elements to encode and recognize the road signs.

Within the Discriminant battery, no measures were significantly correlated with performance on the RSPT.  However, performance on the NAART, Torque Test, and Handedness Inventory did correlate with performance on the RSPT, though not significantly.

Although it was speculated that mood might correlate with performance on the RSPT, based on self-reported mood state this correlation was not supported (BDI: r = .034, $p > .05$), indicating that there is no need to adjust normative data for the presence of depressive symptoms.

Overall, 1/3 of Attentional, 3/3 of Working Memory, and 2/3 of Visuospatial tests were significantly correlated with the RSPT. These findings indicate the RSPT measures a construct related to visual sustained attention, and working memory, as well as some aspects of visuoperception (most obviously simple visual acuity, probably recognition of distinctive elements). In addition, performance on the RSPT does not appear to be affected by the mood of the test-taker, which was acknowledged to be a potential weakness of the test. Correlations between performance on the RSPT, and verbal intelligenceor handedness did not rise to significant levels. Overall, none of the 4 discriminant variables correlated significantly with RSPT performance.

**Further Analyses**

Additional analyses were conducted to assess other factors which could have impacted our sample. In light of the initially poor correlation between the RSPT and the variables selected to establish convergent validity (see Table 1, n = 98 column), two main problem areas were considered: a) there is something unusual about the examiners who administered the tests, or b) there is something unusual about the present sample. Additional considerations were also investigated.

Regarding the inter-rater reliability of the examiners, subjects were coded by their examiner, and between-subjects ANOVAs were conducted using examiner as a nominal IV, and each of the tests administered by the examiners as DVs. The results of the ANOVA indicated there were no significant differences between any of our examiners on the 14 measures

21

completed.  Therefore, our low validity was not be the result of poor inter-rater reliability between our examiners.

There was evidence which suggested that some of our subjects may have given suboptimal effort on the RSPT and other measures.  Subjects were not given any effort tests (as poor effort was not anticipated), but Dr. Green has suggested that the RSPT can be used as an effort test by comparing the present sample's healthy volunteers' performance to that of examinees' with actual cognitive impairment.  To that end, Dr. Green has updated the RSPT to include statistics for each examinee's performance on the 63 easiest vs. the 63 hardest items on the RSPT.  An analysis of the present sample revealed 7 administrations across 6 participants whose performance was more than two standard deviations below the mean for the control group used to establish preliminary data for the RSPT's embedded validity scale.  Two of those participants' performance was more than a standard deviation below the mean for the clinical sample with motivation to malinger who also failed the Word Memory Test.  The 6 participants identified as giving suboptimal performance were excluded from our main analyses, with resulting significant improvements in measures of construct validity (Table 1, Change column). This supports Dr. Green's assertion that the RSPT can also be used to assess client motivation.

There were noteworthy practice effects between test and retest on the RSPT which seemed to be much greater for some participants more than others (Figures 2 & 3).  One reason for this variation may be related to the considerable variance in the latency between test and retest.  The testing was anticipated to take 3 hours, but the actual range of time between test and retest was 38 to 101 mins.  However, there was no correlation for latency time between test-retest and improvement at retest ($r = -.01$), so the generally observed practice effects cannot be wholly attributed to the short duration of the latency interval.

| TEST | | RETEST |
|---:|:---:|:---|
| | 12 | 0 |
| 0 1 1 2 4 5 | 11 | 8 7 6 6 5 5 5 3 3 3 3 2 2 0 0 0 0 |
| 0 0 0 1 1 1 1 1 1 1 1 1 2 2 3 3 3 3 4 4 4 6 7 7 7 9 | 10 | 9 9 9 8 8 8 8 8 8 8 7 7 7 7 7 7 7 7 6 6 5 5 5 5 4 4 3 2 2 2 2 2 2 1 1 1 1 0 0\|4 4 |
| 3\|0 0 0 1 1 1 1 2 2 2 2 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5 5 5 6 7 7 7 7 7 8 8 8 9 9 9 9 | 9 | 9 9 9 9 9 9 8 8 8 7 7 7 6 6 5 5 4 3 3 2 2 0 0 0 0 0 0 |
| 2\|2 2 4 5 5 6 6 6 8 8 8 9 9 9 9 | 8 | 9 9 8 7 6 5 5 2 2\|8 |
| 479 \|57 89 | 7 | \|9 4 |
| 9\| | 6 | \|6 |

Figure 3. Stem-and-Leaf plot of test and retest scores for 98 participants who completed the RSPT with scores excluded from subsequent analyses separated and gray-scaled.

Figure 3 notes: The Stem-and-Leaf Plot, so named for its visual resemblance to a leaf or leaves extending from the stem of a plant, is a relatively simple way to depict large datasets with one or two variables. The graphic gives a visual display of the range and distribution of data points across the range. It provides an interesting way to display every measure of central tendency in one graphic, and does so by using actual datapoints to represent the distribution. The center column or "stem" represents all digits but last, and represents the number of tens in each datapoint. The "leaves," or the numbers to the left and right of the stem, represent the final digit of each datapoint (test scores to the left, and retest scores to the right).

## DISCUSSION

The RSPT is a new screening instrument for driver advisement. The results of test-retest reliability indicate that the RSPT is marginally consistent across multiple administrations. However, this is not surprising as the comparison used raw accuracy performance, rather than standardized scores, to make the comparison. Even measures with high test-retest reliability, such as digit span from the WAIS-III, have lower test-retest reliability for their raw scores (Barr, 2003), and it is reasonable to expect the RSPT's test-retest reliability to improve when standardized scores are prepared. This is important because the utility of the RSPT as a decision-making tool depends on its ability to consistently assess a subject's current functioning. Once normative data is collected, the present data can be reanalyzed using the normative data, to see what improvements standardized norms make to the instrument's temporal stability. Internal consistency as measured by Cronbach's α indicates adequate reliability for time of stimulus presentation and across different signs, but lower reliability of content between trials when the number of stimuli is varied. This is most likely due to a couple of factors: 1) time of stimuli presentation trumps the number of signs in its impact on the difficulty of the trial; 2) the presence of distractors which are not counted towards the number of stimuli presented mean that single target trials with distractors are effectively two stimuli trials, two stimuli trials with distractors are effectively three stimuli trials, and three stimuli trials with distractors are effectively four stimuli trials, blurring the distinction between the variable of number of stimuli presented. This could be adjusted by including data on the distractors in subsequent updates for the RSPT. Overall, these findings indicate that the RSPT is capable of achieving an acceptable degree of reliability on which to confidently base neuropsychological decisions once the final version of the test is completed.

Despite limitations to our analysis of construct validity created by the author's choice in measures of convergent validity, the results of Pearson's r correlations indicate that the RSPT measures working memory, some aspect of attention, and some aspects of visuoperception (most obviously simple visual acuity, probably recognition of distinctive elements).

Performance on the RSPT does not appear to be affected by the mood of the test-taker, which the author hypothesized to be a potential weakness of the test. Correlations between performance on the RSPT, and verbal intelligence or handedness did not rise to significant levels, although verbal intelligence, which correlates with most other domains of cognitive functioning, did correlate more than was expected. Overall, none of the 4 discriminant variables correlated significantly with RSPT performance, indicating that the test possesses a degree of specificity with regards to the cognitive abilities it is measuring.

In addition to the RSPT's potential use as an assessment measure of skills related to driving ability, Dr. Green has also suggested the scale could be used to measure client effort and motivation by detecting suboptimal effort. This assertion was supported, as eliminating participants who were identified as giving suboptimal effort from the present sample improved the correlation between the RSPT and other tests measuring similar constructs.

**Possibilities for Further Research**

In addition to the analyses conducted in the present study, this research has created a database from which additional analyses can be conducted.

While the current study examined subjects' accuracy on the RSPT, the RSPT also contains data on the response time of each participant. A follow-up study should examine response time and accuracy across the different item difficulties of the RSPT, as well as for

individual signs, in order to evaluate whether specific items are consistently missed wherever they appear, as well as what contributions the distractors make to RSPT performance.

Another potential follow-up study, which cannot be examined from the present data, is the possibility that the attention and visuospatial aspects of RSPT might be measuring convergent eye movements or visuo-spatial attention, as described in the "zoom-focus model" of selective attention (Eriksen, & St. James, 1986). Such a hypothesis could be tested by administering the RSPT to participants and monitoring their eye movements through an eye-tracking apparatus.

Finally, performance on the RSPT was significantly correlated with the sex of the participant (e.g., males outperformed females). This may be due in part to the limited number of males in the present sample, and additional male participants should be sampled to balance our numbers before inferences can be made and conclusions drawn regarding sex differences on the RSPT.

**Summary and Conclusions**

The present research provided a preliminary examination of some of the RSPT's operating characteristics. The results indicate that the test has potential to reliably assess working memory, as well as some aspects of attention and visuoperception. This supports Dr. Green's claim that the test can be used to predict impairments related to driving ability. However, the reliability and potential sensitivity of the RSPT is limited for our present analyses by the lack of normative data. Once normative data is completed and the test is ready for commercial use, the present data can be reevaluated to confirm these preliminary claims. In the meantime, the RSPT remains a promising new instrument.

# REFERENCES

Ackerman, M. L., Crowe, M., Vance, D. E., Wadley, V. G., Owsley, C., & Ball K. K. (2011).The Impact of Feedback on Self-rated Driving Ability and Driving Self-regulation Among Older Adults. *The Gerontologist*, *51*(3) pp. 367-378.

Anastasi, A. (1988). *Psychological Testing*, (6th ed.).  NY: Macmillan.

Anstey, K. J., Wood, J. (2011). Chronological age and age-related cognitive deficits are associated with an increase in multiple types of driving errors in late life, *Neuropsychology*, *25*(5), 613-621.

Arditi, A., & Cagenello, R. (1993).  On the Statistical Reliability of Letter-Chart Visual Acuity Measurements. *Investigative Ophthalmology & Visual Science*, *34*(1). 120-129.  retrieved on 3/1/2012 from **http://www.iovs.org/content/34/1/120.full.pdf**.

Barr, W. B. (2003).  Neuropsychological testing of high school athletes. Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology*, *18*, 91-101.

Benton, A. L. (1983).  *Contributions to neuropsychological assessment*, NY: Oxford.

Blanton, P.D., & Gouvier, W.D. (1987).Sex differences in visual information processing following right cerebrovascular accidents. *Neuropsychologia*, *25*, 713-717.

Blau, T. H., (1974). *The torque test: A measurement of cerebral dominance*, NY:APA.

Blau, T. H., (1991).  *The psychological examination of the child*, NY: Wiley.  136-139.

Bondi, M. W., Salmon, D. P., & Kaszniak, A. W. (2009).The Neuropsychology of Dementia. In *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*(Igor Grant & Kenneth M. Adams (Eds.).  159-198. NY: Oxford.

Carr, D. B., & Ott, B. R. (2010).  The older adult driver with cognitive impairment: It's a very frustrating life, *Journal of the American Medical Association*, *303*(16), 1632-1641.

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, *6*(4)**,** 284-290.

Cohen, J (1992). A power primer. *Psychological Bulletin*, *112* (1), 155–159.

Cohen, M.S. (2008).  *Handedness Questionnaire*. Retrieved from **http://www.brainmapping.org/shared/Edinburgh.php#**.

Cooper, P. J. Tallman, K., Tuokko, H., & Beattie, B. L. (1993). Vehicle crash involvement and cognitive deficit in older drivers, *Journal of Safety Research*, *24*(1), 9-17.

Cronbach, L. J., & Meehl, P. E. (1955).  Construct validity in psychological tests. *Psychological Bulletin*, *52*(4). 281-302.

Department for Transport (2009). "Collisions involving older drivers: an in-depth study, "*Road Safety Research Report* 109, London, UK.

Desapriya, E., Wijeratne, H., Subzwari, S., Babul-Wellar, S., Turcotte, K., Rajabali, F., Kinney, J., & Pike, I.  (2011). Vision screening of older drivers for preventing road traffic injuries and fatalities. *Cochrane Database of Systematic Reviews*, *16*(3), retrieved on 3/1/2012 from h**ttp://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006252.pub3/full**.

Dobbs B.M., & Schopflocher D. (2010). The introduction of a new screening tool for the identification of cognitively impaired medically at-risk drivers: The SIMARD A Modification of the DemTect. *Journal of Primary Care and Community Health,* 1(2), 119-127.

Drag, L. L., & Bieliauskas, L. A. (2010).  Contemporary review 2009: Cognitive aging, *Journal of Geriatric Psychiatry and Neurology*, *23*(2) pp. 75-93.

Eberhard, J. (2008). "Older drivers' "high per-mile crash involvement": The implications for licensing authorities," *Traffic Injury Prevention*, *9*(4), 284–290.

Eriksen, C; St. James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model.  *Perception & Psychophysics 40* (4), 225–240**.)**

Epstein, L. J. (1976). Symposium on age differentiation in depressive illness: Depression in the elderly. *Journal of Gerontology*, *31*(3), 278-282.

Evans, L. (2004). *Traffic Safety*, Science Serving Society, Bloomfield Hills, MI, USA.

Green, P. (2008, Oct.). *Road Sign Perception Test*. Paper presented at the Twenty-eighth Annual Conference of the National Academy of Neuropsychologists, New York, NY. Retrieved from **http://www.wordmemorytest.com/rspt.php**.

Groeger, J. A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task*, Hove, UK: Psychology Press.

Hale, P. N., Gouvier, W. D., Schweitzer, J. R., & Shipp, M.  (1987). A small scale vehicle for assessing and training driving skills among the disabled.  *Archives of Physical Medicine and Rehabilitation*, *68*, 741-742.

Hole, G. (2007). *The Psychology of Driving*, LEA, Mahwah, NJ, USA.

Hooper, H. E. (1958). The Hooper Visual Organization Test Manual. Bererly Hills, CA: Western Psychological Services.

Innes, C., Jones, R., Dalrymple-Alford, J., Severinsen, J., & Gray, J. (2009). Prediction of driving ability in people with dementia and non-dementia-related brain disorders. Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Iverson, D. J., Gronseth, G. S., Reger, M. A., Classen, S., Dubinsky, R. M., & Rizzo, M. (2010). Practice parameter update: Evaluation and management of driving risk in dementia, Report of the quality standards subcommittee of the American Academy of Neurology, *Neurology*, *74*(16), 1316-1324.

Justino, L., Kergoat, M., Bergman, H., Chertkow, H., Robillard, A., & Kergoat, H. (2001). Neuroretinal function is normal in early dementia of the Alzheimer type, *Neurobiology of Aging*, *22*, 691-695.

Lee, H. C., Lee, A. H., Cameron, D., & Li-Tsang, C. (2003). Using a driving simulator to identify older drivers at inflated risk of motor vehicle crashes. *Journal of Safety Research*, *34*(4), 453-459.

Lovie-Kitchin, Jan E. (1988). Validity and reliability of visual acuity measurements. *Ophthalmic and Physiological Optics*, *8*(10), pp. 363-370.

Lu, Z. L., Neuse, J., Madigan, S., &Dosher, B. A. (2005). Fast decay of iconic memory in observers with mild cognitive impairments, *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 1797–1802.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*, 341-348.

McGwin, G., & Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers, *Accident Analysis & Prevention*, *31*(3), 181–198.

McGwin, G., Owsley, C., & Ball, K. (1998).Identifying crash involvement among older drivers: agreement between self-report and state records, *Accident Analysis & Prevention*, *30*(6), 781–791.

Mielke, R., Kessler, J., Fink, G., Herholz, K., &Heiss, W.D. (1995). Dysfunction of visual cortex contributes to disturbed processing of visual information in Alzheimer's disease, *International Journal of Neuroscience*, *82*, 1-9.

Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin*, *86*(2). 376-390.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621-640.

Murphy, K. R., Davidshofer, C. O. (2005). *Psychological testing: Principles and applications*, 6th ed. Upper Saddle River, NJ: Pearson.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Pachana, N. A., Fitzell, A. and Lie, D. C. (2004). Clinical and assessment tools to determine driving risk and progress to driving cessation in older adults, Journal of the International Neuropsychological Society: Abstracts Presented at the 2004 Annual Meeting of the Australian Society for the Brain Impairment (ASSBI) and the International Neuropsychological Society (INS). *The 2004 Annual Meeting of the Australian Society for the Brain Impairment (ASSBI) and the International Neuropsychological Society (INS)*, Brisbane, Australia, (23-23). 7-10 July, 2004.

Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and prediction.* Fort Worth: Harcourt.

Reger, M. A., Welsh, R. K., Watson, G. S., Cholerton, B., Baker, L. D., & Craft, S. (2004). The relationship between neuropsychological functioning and driving ability in dementia: A meta-analysis, *Neuropsychology*, *18*(1), 85–93.

Ross, L. A., Browning, C., Luszcz, M. A., Mitchell, P., & Anstey, K. J. (2011). Age-based testing for driver's license renewal: Potential implications for older Australians. *Journal of the American Geriatrics society*, *59*(2), 281-285.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Bury St. Edmunds, England: Thames Valley Test Co.

Salthouse, T. A., & Babcock, R. L. (1991).Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), pp. 763-776.

Schultheis, M. T., & Mourant, R. R. (2001). Virtual Reality and Driving: The road to better assessment of cognitively impaired populations. *Presence: Teleoperators and Virtual Environments*, *10*(4), 436-444.

Schultheis, M. T., Rebimbas, J., Mourant, R., & Millis, S. R. (2007). Examining usability of a virtual reality driving simulator. *Assistive Technology*, *19*, 1-8.

Schultheis, M. T., & Rizzo, M. (2001). The application of virtual reality technology in rehabilitation, *Rehabilitation Psychology*, *46*(3), 296-311.

Schanke, A. K. Sundet, K. (2000). Comprehensive driving assessment: Neuropsychological testing and on-road evaluation of brain-injured patients, *Scandanavian Journal of Psychology*, *41*(2). 113-121.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006).*A Compendium of Neuropsychological*

*Tests: Administration, Norms, and Commentary, 3rd Ed*.  New York: Oxford

Streiner, D. L. (2003).Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Uc, E. Y., Rizzo, M., Johnson, A. M., Emerson, J. L., Liu, D., Mills, E. D., Anderson, S. W., & Dawson, J. D. (2011). Real-life driving outcomes in Parkinson disease, *Neurology*, *76*(22), 1894-1902.

Voyer, D., Voyer, S., Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250-270.

Wechsler, D. (1997).  WMS-III: Wechsler memory scale administration and scoring manual. San Antonio, TX: Psychological Corp.

Whipple, E., Mace, L., Manning, K., Ang, J., & Schultheis, M. (2011, November). *Are Brain-injured individuals at higher risk for simulation sickness in virtual reality?*  Poster session presented at the annual meeting of the National Academy of Neuropsychology, Marco Island, Florida.

**LSU**
LOUISIANA STATE UNIVERSITY

# Consent Form

1. Study Title: Examining reliability and validity of the Road Sign Perception Test

2. Performance Site: Psychological Services Center

3. Investigators: The investigator listed below is available to answer questions about the research,

> M-F, 8:00 a.m. - 4:00 p.m.
> Wm. Drew Gouvier, Ph.D
> 225-578-4138

4. Purpose of the Study: The present study will examine the construct validity, internal consistency, and test-retest reliability of RSPT.

5. Subject Inclusion: Individuals, ages 18-older.

6. Number of Subjects: 150

7. Study Procedures: Participants will be given four brief test batteries (an attention battery, a working memory battery, a visuospatial battery, and a battery of theoretically predicted discriminant variables) and the RSPT twice.

8. Benefits: There are no direct benefits to the subjects. However, information gained from the study will be used to develop the RSPT, which may one day be used to identify the onset of cognitive decline, allowing to professionals begin interventions for dementia sooner.

9. Risks/Discomforts: Since the present research requires that research participants complete a number of psychological tests (estimated time to completion: 2-2.5 hours), the amount of harm or discomfort anticipated in the research is not greater in and of itself than those ordinarily encountered during routine psychological tests. No identifying information will be collected (e.g., names, student ID#s, contact info) except what is required to provide participants with credit for attending and to obtain informed consent, and research assistants will not be permitted to test participants that they know personally. This should minimize social risks such as embarrassment or humiliation if they perform poorly.

10. Right to Refuse: Subjects may choose not to participate or to withdraw from the study at any time.  Subjects who withdraw early for any reason will not be penalized or denied the extra credit earned through their participation.

11. Privacy: The LSU Institutional Review Board (which oversees university research with human subjects) may inspect and/or copy the study records.  Results of the study may be published, but no names or identifying information will be included in the publication.

   Other than as set forth above, subject identity will remain confidential unless disclosure is legally compelled.

12. Financial Information: There is no cost to the subjects.  Subjects will be compensated for participating in the study with extra credit in their undergraduate psychology class.

Signatures:

The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigators. If I have questions about subjects' rights or other concerns, I can contact Robert C. Mathews, Institutional Review Board, (225) 578-8692, irb@lsu.edu, www.lsu.edu/irb. I agree to participate in the study described above and acknowledge the investigator's obligation to provide me with a signed copy of the consent form.

Subject Signature:_____ Date: _____

The study subject has indicated to me that he/she is unable to read. I certify that I have read this consent form to the subject and explained that by completing the signature line above, the subject has agreed to participate.

Signature of Reader: _____ Date: _____

Institutional Review Board
Dr. Robert Mathews, Chair
203 B-1 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu | lsu.edu/irb

Psychological Services Center • Louisiana State University • Baton Rouge, LA  70803-5501 • 225-578-1494 (P)

# VITA

Justin H. Ory was born and raised in southern Louisiana. He graduated from Mandeville High School in 1999, and from Southeastern Louisiana University with Bachelor of Arts and Master of Arts degrees in 2005 and 2008 respectively. As an undergraduate student at Southeastern Louisiana University, Mr. Ory had the pleasure of working with Hunter A. McAllister, Ph.D., examining methods of improving the reliability of eyewitness identification in line-ups or mugbooks. As a graduate student at Southeastern Louisiana University, Mr. Ory continued his work with Dr. McAllister, and completed a master's thesis in evolutionary psychology with Matthew Rossano, Ph.D. He enrolled in the clinical psychology doctoral program at Louisiana State University under the mentorship of Wm. Drew Gouvier, Ph.D., in 2009. His research interest is focused on psychological tests and measures, particularly with test construction and development. His clinical interests include psychological testing and diagnosis of learning disorders, ADHD, and neurocognitive disorders in adolescents and adults.