INDIVIDUAL DIFFERENCES AND IN-VEHICLE DISTRACTION WHILE DRIVING: A TEST TRACK STUDY AND PSYCHOMETRIC EVALUATION

Louis Tijerina
Transportation Research Center Inc
East Liberty, OH 43319

Edwin B. Parmer
Transportation Research Center Inc
East Liberty, OH 43319

Michael J Goodman
National Highway Traffic Safety Administration
Washington, DC 20590

ABSTRACT

The influence of individual differences on driver distraction was examined in this study. Sixteen (16) test participants were trained on destination entry procedures with four commercially available route guidance systems, as well as the dialing task on a commercially available wireless cellular telephone and on manually tuning an after-market car radio. The participants then drove an instrumented vehicle at approximately 45 mph on a 7.5 mile oval test track with very light traffic while concurrently engaging in various tasks with these devices. In-vehicle task completion time, average glance duration away from the road ahead, number of glances away from the road ahead, and number of lane exceedences were recorded. The participants were later given an automated battery of temporal visual perception and cognitive tasks. Performance on the test battery was then correlated to performance on the test track measures to determine the extent to which individual driver differences could account for observed performance differences. Analysis of these elementary test scores as predictors show low but consistent patterns of correlation to test-track performance measures.

INTRODUCTION

The proliferation of information and telecommunications systems for use in cars and trucks has made driver distraction a pressing highway safety concern. Driver distraction or workload reflects three major influences: the nature of the in-vehicle device or task, the driving conditions under which that task is pursued, and the individual abilities of the driver. Existing research has focused primarily on assessing driver distraction in terms of the in-vehicle device or task (e.g., Tijerina, Parmer, and Goodman, 1998). Less attention has been devoted to characterizing the driving conditions under which a task might be pursued (e.g., Hulse, Dingus, Fischer, and Weirwille, 1989). Even less research has been conducted to assess how individual differences among drivers might influence their propensity toward distraction.

Research into individual differences and distraction might contribute to highway safety in several ways. For commercial vehicle operations, human abilities tests might be identified that correlate substantially with safe driving. Kahneman, Ben-Ishai, and Notan (1973), for example, found performance on an auditory shadowing task to be significantly correlated with crash involvement among a sample of Israeli truck drivers. Identification of human abilities associated with time-sharing skill in a driving context might lead to new methods of driver training. Individual differences research can also support system design. For example, drivers differ in their spatial abilities as measured by psychometric tests. Such differences manifest themselves when drivers must use moving map displays in route guidance systems. On the other hand, drivers make good use of egocentrically-defined text directions regardless of their spatial abilities (McGehee, personal communication).

Recent research has identified a set of temporal factors in visual perception and cognitive factors might be predictive of real world performances (Kennedy, et al.,1997). It was desired to determine the extent to which such tests, provided in a computerized battery termed PATSYS, might be indicative of performance during in-vehicle device use while driving.

METHOD

Test Participants: Sixteen (16) Transportation Research Center Inc. test drivers participated. These drivers were hourly employees with valid driver's licences and generally less than 2 years of TRC driving experience. There were equal numbers of males and females in each of two age categories: Younger (35 years or younger) and Older (55 years or older). None of the test participants owned or had significant prior experience with route guidance systems or cellular telephones prior to this study.

Test Vehicle and Instrumentation: The test vehicle was a 1993 Toyota Camry, equipped with Micro-DAS instrumentation (Barickman, 1998) which captured travel speed, lane position, and lane exceedences, as well as video of the road scene and driver eye glance behavior at a 30-Hz sampling rate. Eye glance video was later manually reduced.

Route Guidance Systems: Four (4) unmodified, commercially available route guidance systems. The dash mounted Delco Telepath 100® consisted of a 3-line LCD display to present menu items, scrolled by means of a bezel-mounted rotary knob and selected by pressing an Enter key. The Alpine NVA-N751A® incorporated a free-mounted 5.6 inch active matrix color display without bezel keys. It displayed an alphanumeric keyboard and entries were made by scrolling from key to key with a joystick mounted on a remote control unit; pressing down on the joystick registered a character or selection. If sufficient alphanumerics were entered for the system to estimate candidate destinations, these were presented as an alphabetized scrolling list of 3 items at the bottom of the display of the alphanumeric keyboard screen. The Zexel Navmate® consisted of a free-mounted 4 inch diagonal full color LCD screen with a set of bezel control keys, including a central "left, right, up, down" key and an Enter key. Both the Zexel and Alpine systems were mounted on a gooseneck pedestal bolted to the floor board between the driver and passenger. The Zexel system presents menu options for destination entry type and city, followed by a scrolling display of numerically and alphabetically arranged destinations generally presented 11 to13 lines at a time. The driver presses the Enter key to make a selection. Finally, the dash mounted Clarion Eclipse® Voice Activated Audio Navigation (VAAN) system used voice recognition and output exclusively; there was no visual display. Keywords would activate the VAAN for destination entry. Destinations were entered by spelling them. The VAAN emphasized precise spelling of a destination; each letter uttered by the driver would be proceeded by a beep to acknowledge receipt of the input. The driver uttered "verify" to conclude an entry. The system would eventuate in a spoken list of best-guess candidate destinations for selection by the driver via YES or NO verbal responses.

The last three of these systems allowed for entry of a street address, intersection, or point of interest (attraction, restaurant, hotel, etc.). Thus, three types of tasks (address, intersection, point of interest) were included as suitable for comparisons among the systems. The Delco system only supported point of interest selection. Also, two additional tasks were included for comparison purposes: tuning a radio to a specific band and frequency with a modern "Seek" function on the Clarion Eclipse system; or manually dialing a cellular telephone (a 10-digit number on a handwritten note card) using a cordless AUDIOVOX Model MVX-500.

The PATSYS test battery was used to conduct the psychometric evaluation (See Table 1). This battery was run on a Gateway2000 E-3110 personal computer with a Vivitron 17-inch diagonal high resolution color monitor. For further details of these tests, see Kennedy, Silver, and Ritter (1995), Turnage and Kennedy (1995), and Kennedy, Turnage, and Lane (1997).

<u>Test Track</u>: The Transportation Research Center Inc. (TRC) 7.5-mile multi-lane test track is in the form of an oval with banked curves at either end and with unbanked straightaways that measure approximately 2.0 miles each. The test track is comprised of three 12-ft wide concrete lanes with a fourth inner blacktop lane for use in the event of vehicle breakdowns or required stops. The test vehicle for this study operated in lane 1 (adjacent to the innermost blacktop lane) and changed lanes only as needed for normal track operations and safety. The test participant was asked to drive at approximately 45 mph on the straightaways and accelerate to 60 mph on the curves, provided that any requested tasks are completed by the time the test vehicle enters a curve. Otherwise, the driver was to maintain 45 mph and attempt to complete the requested in-vehicle task. Traffic density tended to be light relative to open road driving. However, travel speeds for other vehicles of the track might vary greatly, vehicles involved with other testing could slow, stop, or move to the blacktop lane abruptly, and track repair and roadside obstructions had to be avoided. Faster traffic drove on the outer lanes of the oval. Data collection was scheduled for between 8:00 am and 4:30pm weekdays.

Procedure: Prior to the data collection runs, the experimenter familiarized the test participant with each navigation system. Each test participant then completed 12 practice data entry tasks per system (four for each destination category), entered while the vehicle was parked. This training was done in two phases (morning and afternoon); so, two systems were reviewed prior to each half of the test track trials. On the 7.5 mile track, the order of trials were counterbalanced across the four route guidance systems (Zexel, Alpine, Delco, and VAAN), destination entry category (point of interest, intersection, and street name targets), and target (Target A or Target B within a category). All trials with a given system were executed before moving on to another system; the destination type and targets within destination type were counterbalanced to control for order effects. The cellular phone and radio tuning tasks were interspersed between destination entry trials on an opportunistic basis by the experimenter in a quasi-random fashion. Prior to leaving for the test track, the destinations were presented to the test participant in 18-point Times Roman font and the test participant was asked to write in his or her own hand each destination on a separate index card, as well as the 10-digit unfamiliar telephone number, such that they would be able to read from it while driving. A task began when the ride-along experimenter gave the driver a hand-written card or a radio tuning task was requested orally by the ride-along experimenter. The task ended when the request had been fulfilled, as indicated by an event marker triggered by the experimenter. Requests for tasks were generally made when the test participant was exiting a curve onto a straightaway segment of the test track. After test track data collection was completed, the test participant answered the subjective assessment questions and was released. Each test participant was invited back subsequent to the test track trials and administered the battery of temporal acuity and cognitive tests. The battery of tests was administered four times in a single day. The results from the last of the four rounds of testing were used for data analysis.

Test Track Measures, Test Battery Measures: Four response measures from the test track study were selected for analysis: in-vehicle task completion time per trial (TASKTIME, seconds), mean average glance duration to the device during task completion (MNGLNCTM, seconds), glance frequency or number of glances to a given in-vehicle device per trial (GLNCFREQ), and number of lane exceedences or departures per trial or task completion (NEXCEED). The previously mentioned test battery subtests generated latencies for all TEMPORAL tests (see Kennedy, et al, 1995 for explanation of these) and for the RT4 test; all other cognitive tests were scored in terms of number of trials correct over a fixed period of testing (not a fixed number of trials).

RESULTS

The data were analyzed in terms of correlation and regression. Table 2 shows the matrix of intercorrelations among test track and test battery measures. This table reveals that among the test track measures a) task time is highly correlated with glance frequency to the device, b) both are moderately correlated with number of lane exceedences (NEXCEED), and

c) mean glance duration is not correlated with any of the other test track measures. This pattern of results is comparable and consistent with other studies examining the effects of in-vehicle device use while driving (Green, 1998).

Among the battery of temporal and cognitive tests there is a moderate degree of intercorrelation among the temporal tests and high intercorrelations among the cognitive tests. However, the intercorrelations across the two subsets of tests are generally lower than within each subset, indicating that they reflect distinct aspects of human performance.

An all-possible-regressions analysis was carried out using the PROC REG procedures in SAS. For each of the test track measures, all possible models were assessed for the various combinations of those test battery tests with a statistically significant correlation ($\alpha \leq 0.05$). The criterion used to select the "best" model was the adjusted R^2 criterion. The adjusted R^2 criterion is equivalent to finding the set of predictor variables that minimizes the residual mean square error for the model (Montgomery and Peck, 1992).

In Table 2 reasonable patterns of correlation appear with task completion time. TASKTIME worsens (i.e., increases) as DVA, STR, MSK, and RT4 scores worsen (i.e, increase) and TASKTIME improves (i.e., decreases as CS, PC, GR, and MNK scores improve (i.e., increase). The only anomaly is PHI; TASKTIME decreases as PHI scores worsen (i.e., increase). Overall, as measures of temporal acuity, perceptual speed, working memory, speed of processing, spatial abilities, and higher cognitive processes improve, TASKTIME decreases reliably. However, the proportion in TASKTIME variability that covaries with a set of such regressors is modest. The "best" subset of regressors for TASKTIME were PHI, RT4, and GR with Multiple R = 0.35, Adjusted $R^2 = 0.12$.

Correlations between significant PATSYS tests and glance frequency (GLNCFREQ) also are consistent with intuition. Glance frequency worsens (i.e., increases) as DVA, STR, MSK, and RT4 scores worsen while GLNCFREQ improves (i.e., decreases) as CS, PC, GR, and MNK scores improve. However, the proportion of variability in GLNCEFREQ that covaries with the "best" regression model is only about 10 percent. The "best" set of predictors for GLNCFREQ were STR MSK, and PC, with Multiple R = .33, Adjusted $R^2 = 0.10$.

The only variable selected as a regressor for mean glance time (MNGLNCTM) was MSK, with $r^2 = 0.03$. As masking scores worsened (i.e., increased), so did mean glance time.

Correlations between significantly correlated PATSYS tests and the incidence of lane exceedences (NEXCEED) were all sensible in sign. NEXCEED measures worsened (i.e., exceedences increased) as DVA, STR, MSK, and RT4 scores worsened (i.e., increased). NEXCEED measures improved (i.e., decreased) as CS, PC, and MNK scores improved. The best subset of regressors identified were MSK, PC, and MNK, with a multiple R = 0.44, adjusted $R^2 = 0.19$.

DISCUSSION

This study represents an attempt to assess the explanatory power of individual differences in both temporal acuity and cognitive abilities in terms of various measures of driver distraction or workload while using a variety of in-vehicle devices. The variability shared in common between a given measure of test track performance and the "best" subset of test battery measures is modest at best. This perhaps reflects the relative contribution of individual differences (as measured by these tests) to in-vehicle task completion while driving. This finding is consistent with other research into individual differences and highway safety (Elander, West, and French,1993). It would not be surprising to find that the specifics of the task and driving conditions at the time of task execution, combined with driver motivation, fatigue, and the like command a much larger share of the variability in task outcomes. There is also random errors that arise in device use and a variation in error recovery that also increase response variability.

When each dependent measure was examined within the context of specific test battery components, there was high face validity to predictor sets. Thus, better task time was associated with better temporal acuity, faster processing and higher cognitive capabilities. Likewise, reduced glance frequency was associated with better dynamic visual and temporal acuity, better pattern comparison performance and faster processing of information. These relationships and degree of overlap suggests that with greater refinement, efficiency and packaging of the test battery it may be possible to tune in-vehicle tasks to the specific cognitive and temporal capabilities of individual drivers, a step towards building truly "intelligent" systems. Future work should examine such refinements and explore the more subtle relationships between specific task demands and predictor sets.

REFERENCES

Elander, J., West, R., and French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. *Psychological Bulletin*, *113*(2), 279-294.

Green, P. (1998, April). *Visual and task demands of driver information systems*. Internal working draft. Ann Arbor: University of Michigan Transportation Research Institute.

Hulse, M. C., Dingus, T. A., Fischer, T., and Wierwille, W. W. (1989). The influence of roadway parameters on driver perception of attentional demand. In A. Mital (Ed.), *Advances in industrial ergonomics and safety* I (pp. 451-456). London: Taylor and Francis.

Kahneman, E., Ben-Ishai, R., and Lotan, M. (1973). Relation of a test of attention to road accidents. *Journal of Applied Psychology*, *58*, 113-115.

Kennedy, R. S., Silver, N. C., and Ritter, A. D. (1995). A visual test battery: Tale of two computers. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, 907 - 911.

Kennedy, R. S., Lane, N. E., Turnage, J. J., and Harm, D. L. (1997). Individual differences in visual, motor, and cognitive performances: Correlations with a simulated shuttle landing task. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, 589- 593.

Montgomery, D., and Peck, E . (1992). *Introduction to linear regression analysis (2nd edition)*. New York: John Wiley.

Tijerina, L., Parmer, E., and Goodman, M. J. (1998). Driver workload assessment of route guidance system destination entry while driving: A test track study. *Proceedings of the 5th World Congress on Intelligent Transport Systems*, 12-16 October 1998, Seoul, Korea.(CD-ROM).

Turnage, J. J., and Kennedy, R. S. (1995). A hierarchical view of human performance: The neglected contribution of visual factors to predicting perfomrance. In A. C. Bittner and P. C. Champney (Eds.), *Advances in industrial ergonomics and safety VII* (pp. 195-202). London: Taylor and Francis.

Wickens, C. D. (1992). *Engineering psychology and human performance (2nd Edition)*. New York: Haper Collins.

Table 1. PATSYS Test Battery: Temporal and Cognitive Subtests.

TEST NAME	**TEST DESCRIPTION**	**SCORING**
Dynamic Visual Acuity (DVA):	This test varied the presentation time between the letter "C" presented on the left of the video screen and a letter "C" presented on the right of the video screen. The participant's task was to determine if the C's were facing in the same or opposite direction. DVA refers to the ability of an observer to resolve fine detail in an object when there is relative motion between them. This may be important as a driver repeatedly shifts gaze from the road scene to inside the vehicle.	The participant's score was the fastest presentation time for correct responses. A lower score indicates better dynamic visual acuity.
Simultaneity (SIMU):	This test presented two open boxes 33 mm apart which were alternately flashed on the screen for 60 msec. The interstimulus interval (ISI) for onset of the two boxes was manipulated. The basis for this test is that relatively large temporal differences are needed before an observer can reliably perceive two stimuli as non-simultaneous. Reflects visual processing speed, acuity.	The participant's score was the lowest time value or ISI when the two boxes appeared to be on simultaneously. A lower score indicates better temporal acuity.
Bistable Stroboscopic Motion (STR)	This test presented an array of boxes that were alternately cycled. Frame one consisted of three horizontal elements of boxes with equal center-to-center distances. Frame two had identical elements had identical elements shifted to the right by a distance equal to the center-to-center separation between stimuli. Participants responded by keyboard presses whether they perceived "element" motion (appearing as four boxes) or "group" motion (appearing as a set of three boxes that alternatively shifted back and forth laterally one box-width).	The point of transition from one type of motion to the other was collected as a threshold value. A lower score represents greater sensitivity.
Phi Phenomenon (PHI)	Square boxes, 33 mm apart on the video display were presented to the left and right of a fixation point. Through a set of response keys, the participant would adjust the interstimulus interval (ISI) to the point where the boxes appeared to transition from moving successively to a single box moving back and forth. Reflects visual processing speed.	The value of the ISI at the fifth reversal was the participant's score. A lower score would signify better temporal acuity.
Masking Test (MSK)	In this test, two vertical lines .075" in length and 0.05" in width were presented. A horizontal line 0.05" in length extended from the midpoint of either the left or right vertical line. After a brief period, the lines were replaced by a complex pattern of dots (the mask). The screen went blank and the participant was instructed to press the left or right arrow keys depending on whether the horizontal line was on the left or right vertical line. Masking is the interference in the perception of one briefly presented stimulus by a second, succeeding stimulus briefly presented nearby in time and space. This may be important as the driver attempts to retain information in working memory while glancing between road scene and in-vehicle device.	The ISI between the target and the mask was varied and the participant's score was the lowest ISI for correct responses. A lower score signified better temporal acuity.
Grammatical Reasoning (GR)	This test employs five grammatical transformations on statements about the relationship between two letters "A" and B" For example, There are 32 possible items arranged in random order. The participant assesses the correctness of the statement by pressing the "T" key for true statements or the "F" key for false statements. This measures higher cognitive processes of deductive reasoning. It may reflect a general effect of greater cognitive capability on task completion.	The participant's score is the number correct out of 32 statements. A higher score is indicative of greater capacity of higher cognitive processing.
Mannikin Test (MNK)	A simulated human figure (a sailor) is presented in either full-front or full-back orientation on the screen. The figure is shown holding three hearts, diamonds, clubs, or spades, different patterns in each raised hand. One of the two patterns held matches a pattern which appears on a podium the figure stands on. The participant indicates which hand is holding the pattern matching that in the podium by pressing the appropriate key. This test appears to measure ability in mental rotation and related transformations. MNK scores might reflect the ability of a driver to reorient spatially between glances to an in-vehicle display layout and the road scene.	The participant's score is the number correct out of 16 trials. A higher score signifies better spatial ability.
4-Choice Reaction Time (RT4)	On this test, four outlined boxes are displayed above the numbers 1, 2, 3, and 4. At random intervals, one of the boxes illuminates, i.e, changes from outline to filled. The participant presses a corresponding key as quickly an accurately as possible. This test assesses the participant's speed of information processing to make a response from multiple alternative stimuli, depending on which alternative is signaled. Speed of cognitive processing is ubiquitous as an contributor to cognitive task performance.	The participant's score is the latency between when a box illuminates on the screen and when the corresponding key is pressed on the keyboard. Shorter reaction times generally represent faster processing
Pattern Comparison (PC)	In this test, a pair of eight-dot patterns are presented and the test participant indicates on the keyboard whether the two patterns are the same or are different. This is a test of perceptual speed. This may be important as driver's compare entry data to device display feedback.	The participant's score is the number of pairs correctly identified as similar or different. Higher scores imply greater perceptual speed.
Code Substitution (CS)	This test involves a display of nine characters on the top of the screen and beneath them the numbers 1 through 9 in parentheses. Under the code are two rows of characters with empty parentheses beneath them. The participant inserts the number associated with the character from the code displayed at the top of the screen. This test appears to assess working memory and perceptual speed. It may reflect the capability of a driver to keep track of task components.	The score is made up of the number of correctly matched digits to their corresponding letters. Higher scores imply greater perceptual speed and working memory.

Table 1. Intercorrelation Matrix for Test Track and Test Battery Measures.

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0

	TASK TIME	GLNC FREQ	NEX CEED*	MNGL NCTM	DVA	SIMU	PHI	STR	MSK	CS	PC	RT4	GR	MNK
TASK TIME	1.0000 0.0	0.8587 0.0001	0.3797 0.0001	-0.0485 0.3533	0.1995 0.0001	-0.0177 0.7346	-0.1420 0.0064	0.1766 0.0007	0.2241 0.0001	-0.2823 0.0001	-0.3021 0.0001	0.3055 0.0001	-0.2609 0.0001	-0.2932 0.0001
GLNC FREQ		1.0000 0.0	0.2972 0.0001	-0.0688 0.1879	0.1975 0.0001	-0.0173 0.7413	-0.0924 0.0767	0.2420 0.0001	0.1319 0.0113	-0.2358 0.0001	-0.2836 0.0001	0.2699 0.0001	-0.2629 0.0001	-0.2129 0.0001
NEX CEED*			1.0000 0.0	0.0092 0.8630	0.2548 0.0001	-0.0526 0.3242	-0.0759 0.1545	0.1086 0.0414	0.4128 0.0001	-0.2266 0.0001	-0.1764 0.0009	0.2654 0.0001	-0.0601 0.2598	-0.2966 0.0001
MNGL NCTM				1.0000 0.0	0.1349 0.0096	-0.0317 0.5447	-0.0359 0.4919	0.0446 0.3936	0.1731 0.0009	-0.0780 0.1355	-0.0301 0.5649	0.069 0.1842	0.0201 0.7004	-0.1025 0.0494
DVA					1.0000 0.0	-0.0995 0.0564	-0.1633 0.0017	0.4981 0.0001	0.5762 0.0001	-0.4655 0.0001	-0.4251 0.0001	0.4669 0.0001	-0.2753 0.0001	-0.3827 0.0001
SIMU						1.0000 0.0	0.0533 0.3083	-0.0696 0.1824	-0.1565 0.0026	0.0371 0.4779	-0.0584 0.2641	0.2107 0.0001	-0.0130 0.8029	0.2262 0.0001
PHI							1.0000 0.0	-0.0618 0.2367	-0.2217 0.0 001	0.1981 0.0001	0.2448 0.0001	-0.2004 0.0001	-0.0977 0.0611	0.2700 0.00 01
STR								1.0000 0.0	0.4245 0.0001	-0.2676 0.0001	-0.4204 0.0001	0.6460 0.0001	-0.3600 0.0001	-0.1956 0.0002
MSK									1.0000 0.0	-0.6591 0.0001	-0.6177 0.0001	0.6849 0.0001	-0.2829 0.0001	-0.6757 0.0001
CS										1.0000 0.0	0.8497 0.0001	-0.5949 0.0001	0.6392 0.0001	0.7993 0.0001
PC											1.0000 0.0	-0.7535 0.0001	0.6474 0.0001	0.7873 0.0001
RT4												1.0000 0.0	-0.4802 0.0001	-0.6480 0.0001
GR													1.0000 0.0	0.5884 0.0001
MNK														1.0000 0.0

* NOTE: Number of Observations = 368 except NEXCEED = 353.
TASKTIME: In-Vehicle Task Completion Time; GLNCFREQ: Number of glances to device to complete task; NEXCEED: Number of lane exceedences during task completion; MNGLNCTM: Mean glance time to device during task completion.