Correlational Data that Support a Constructive Assessment of Driving Skills

Kenneth C. Mills
Profile Associates
Chapel Hill, NC

Robert C. Hubal
Research Triangle Institute
RTP, NC

ABSTRACT

We describe how a constructive PC-based driver assessment part-task trainer (PTT) can be integrated into driver training. The PTT, developed through research with law enforcement agencies, gathers data on drivers' scanning and divided-attention skills and also measures the tendency toward tunnel vision under stress. The user sits in front of a computer monitor and interacts with the system through a force-feedback steering wheel and foot pedals.

A validation study tested 50 North Carolina State Highway Patrol cadets before they were evaluated on a closed circuit driving course. A composite score was derived that reflected the cadets' scanning and divided-attention skills at five levels of increasing difficulty. The track testing evaluated behavioral skills as the cadet completed a 1.5-mile circuit with 11 obstacles; the skills were decision-making (evasive actions with obstructed visual fields), accuracy of maneuvers (number of cones hit), and lap times on three consecutive laps. Consistent with previous research, the PTT scores showed a linear degradation of skills over the five difficulty levels. Cadets with higher computer scores showed fewer driving errors on the track, especially on lap 2. The evasive action exercises were the most sensitive to individual differences. A replication with 50 additional cadets found a relationship between PTT scores and track instructor ratings of driving skills. The results show that a complex and unfamiliar computerized training assessment correlates with real-world driving skills, particularly when the driver is under pressure.

A survey of trainers employing the PTT suggests improvements in how the system should be integrated into driver training. Results from an evaluation of its effectiveness as an assessment tool point to methodologies for focusing costly on-track (live) training by using relatively low-cost video reality (constructive) assessment data. Implications for developing computerized and track assessments of driving skills are discussed.

Biographies

Dr. Mills researches computerized assessments of visual and cognitive skills. He has conducted clinical trials on the effects of alcohol, stimulants, and fatigue on the time-series enhancement and degradation of cognitive skills. His recent research has explored the development of tunnel vision as a risk factor for driving and piloting in both laboratory and track settings.

Dr. Hubal conducts research on technology assisted learning, focusing on development, presentation, and evaluation of materials and identifying approaches to improve learning and training effectiveness. He has integrated evaluative capabilities into VR-based advanced learning environments, ensuring that systems perform as expected in assisting learning. Most recently he has developed behavioral software that enables virtual humans to act and behave realistically in controlled learning contexts. He has applied research results to such everyday domains as medical informed consent, consumer decision-making, and language learning.

Correlational Data that Support a Constructive Assessment of Driving Skills

Kenneth C. Mills
Profile Associates
Chapel Hill, NC
profiler@intrex.net

Robert C. Hubal
Research Triangle Institute
RTP, NC
rhubal@rti.org

INTRODUCTION

Skills acquisition and practice of any kind in a live environment can be time-consuming, costly, and/or dangerous [17]. In particular, skills acquisition and practice of rarely experienced real-world events in a live environment is at best unsystematic. For perceptual skills, including those used during rarely experienced events, training and assessment can be a haphazard process; even when live training events are staged, formal assessment of mission-critical cognitive skills remains difficult [14]. Fortunately, constructive learning environments have the potential to measure complex, integrated perceptual and cognitive skills, even for rarely experienced events, and give the student contextual guidance and immediate feedback based upon performance. Importantly, dramatic improvements in PC-based tools with high levels of interactivity and fidelity have made part-task trainers (PTT's) available for complex cognitive skill development. This paper reports preliminary data on the validation of a PTT to assess and train scanning and multitasking skills critical for driving.

Driving

Driving is a complex activity involving skills acquisition at varying rates, and at varying levels of control. Like most forms of expertise [5], driving skills range from simple knowledge-based rules to well-integrated and complicated visual/psychomotor responses. Most common models describe driving comprising at least three levels: operational (declarative), procedural, and strategic (see [21,31]). Similar models from other areas of research describe outputs of cognitive task analyses (i.e., of the skills needed to perform a task): factual knowledge; prescribed behaviors (those that have set procedures); and adaptive behaviors (those that require strategic application of knowledge) [6,7]. The process of decomposing learning into separate components causes instructional designers (and students) to better understand what needs to be learned to determine how to learn it.

A variety of studies indicate that becoming adept at visual search and hazard detection (identifying and responding to salient information in a constantly changing scene) happens at both the procedural and strategic levels [31]. The beginning driver may learn to spot important road signs in a weekend ("importance" being declarative knowledge), yet take years to scan and respond to all of the intricate hazards accompanying a left turn at a busy intersection. Inexperienced drivers show deficiencies in their ability to scan ahead, anticipate distant hazards, and integrate different skills so that they can perform more than one task at a time (attention allocation) [19,21,23]. More importantly, eye-movement and accident studies suggest these skills are learned slowly over not months but years [19,22,28,29].

Visual search and hazard detection may be acquired slowly because real-world targets differ widely in familiarity, predictability of location, and movement. Opportunities for systematic practice are limited. Unless costly driving scenarios are staged on a track or test course, conspicuous hazards may occur too infrequently for efficient training. Thus, in the real world, active, goal-directed scanning varies widely among individuals, situation familiarity, and experience [8]. Inexperienced drivers often do not have convenient or readily accessible means to improve these skills. Whereas experienced and professional drivers carry out complex emergency maneuvers smoothly, and with a high level of automaticity, beginning drivers can not. Apparently, integrating complex skill sets in high demand situations requires a fair amount of repetitive practice (see also [12]).

Professional, police, and emergency medical services (EMS) drivers with years of on-the-road experience are generally observed to stay calm in high-demand situations. However, the situation is reversed for the beginning driver, for whom more errors occur as task demands increase, time pressures increase, and decisions are required beyond experience or trained levels [10,19,23]. Novel, unexpected, or high-demand situations for which no applicable rules or skill base is available disrupt skill-based (automatic) processing [10]. While experts and novices base judgments on similar amounts of information, they differ in types of information used. Experts distinguish, and process, more "diagnostic" information [32].

To make matter worse for the beginning driver, high-demand situations can increase the chances of sympathetic arousal and concomitantly occurring tunnel vision [27]. Tunneling appears to be a fairly primitive response to even mildly surprising events and happens in a wide range of training situations [9,12,24]. Tunneling cascades in seconds as the driver narrows his/her focus exclusively on the threat. Gathering extraneous or peripheral information is severely inhibited. It can be particularly devastating while driving because hazards continue to appear from all parts of the visual scene. Tunneling is well documented in youth and elderly drivers [2,30,33]. A constructive learning environment that could increase visual and perceptual demands more frequently than encountered in the real world might further minimize the chances of tunneling [24].

PC Appended Devices and PTT's

Traditionally, procedural and strategic skills development, such as learning to manage task overload, has been restricted to live or highly realistic simulation environments (e.g., Army experiments, flight simulators). But new technologies that enliven personal computers, such as virtual reality (VR), agents, and natural language processing, increase the possibilities for using the PC for skills training and assessment. For instance, a PC-based VR system was found to be more effective than sketches and also an immersive system for aircraft maintenance skills [4]. VR has proven particularly useful to train repair and maintenance skills for a space mission [20] and diagnostics on an Abrams tank [15]. Similarly, a series of applications are geared toward training medical students to learn to insert invasive tools such as catheters and scopes (e.g., bronchioscope, endoscope) [1]. Finally, PC-based responsive virtual humans (agents) are being used to train interaction skills for patient interaction [18].

The commonality among these examples is their use of PC technology and appended devices (e.g., joysticks, microphones) that constitute what we term PTT's. Thus, on a subsequent Abrams maintenance training application, soldiers were given the ability to manipulate the virtual environment via an integrated commanders handle; the medical applications described above use virtual needles and tubes; and the responsive agents converse with the user through natural dialog. Hence, for many training applications, a PC with standard-issue or low-cost hardware attachments, rather than specially developed (and expensive) components, enables training and assessment on procedural and strategic skills.

Safety experts have set out criteria for PTT's with the goal of training higher order perceptual and cognitive skills necessary for good driving [11]. The ideal characteristics of a PTT that would qualify for advanced perceptual training with inexperienced drivers include task variation, task repetition, gradually increasing time and task demands, and scenarios that are not predictable. Theoretically, a PTT presenting multiple fast-paced scenarios could teach inexperienced drivers to anticipate hazards and acquire complex survival strategies beyond those which are limited by single response setups or simple hardware mockups. The introduction of appended PC interfaces (specifically a steering wheel and pedals), coupled with higher processing speeds and more detailed graphics, provide exactly the platform needed for complex, highly interactive scenarios for driver assessment and training. What remains is a means of presenting realistic, dynamic scenarios for the student driver in which to practice. We describe one such application next, a multi-task paradigm with increasing workloads, speeded response requirements, dynamic moving scenes with numerous distractions, a high level of unpredictability, and anticipation required for responses.

DRIVER TRAINING APPLICATION

Basis in Driver Training Research

Whether complex "game-like" scenarios assess skills that are specifically related to real-world driving skills remains open to empirical testing. The following studies examined preliminary data from two PC-based tasks (one simple, one complex, that both assess visual/scanning and divided-attention skills) against intricate closed-course driver evaluation. The simple behavioral task (Performance On-Line, or POL) uses keypress responses to single target and divided-attention stimuli at three discrete stimulus rings outward from the center of a PC display. The task assesses tunnel vision by comparing targets at the three levels of eccentricity. In a previous study with POL, alcohol-induced tunneling was characterized by greater impairment at the outer rings of the display [25]. A subsequent study used the POL task to evaluate police cadets and found a significant correlation between divided-attention scores at the outermost ring on POL and ratings of track driving performance during pursuit training

[26]. The tendency to tunnel, as measured by POL, apparently carries over to performance during emotionally-charged driver training.

The complex PTT (Profiler) uses the scoring algorithm from POL and adds a commercial-off-the-shelf steering wheel and foot pedals package. The interactive software application records driving performance as the driver maneuvers through realistic streets in a PC-based production. It gathers data on drivers' scanning and divided-attention skills, as well as steering and collision avoidance, during a 30-minute test that increases task demands over five levels of difficulty. Current versions offer choices of law enforcement, ambulance, and fire vehicles.

Basis in Human Factors

Profiler uses a technology called Video Reality® that seamlessly integrates characters, sets, and roadways filmed with standard production techniques into a virtual environment. Users enjoy continuous freedom of movement with a 180º field of vision throughout the environment, letting them control where they go, when, and what they look at on the way there. The technology is quite cost-effective for creating a virtual environment for applications wherein the need for photographic realism is paramount and the computing base is limited to standard PC's. It is particularly applicable to part-task driving simulation since (compared with a high-resolution VR model) the virtual environment has the same user-acceptance as a movie and leads to suspension of disbelief and, consequently, sympathetic arousal.

Our studies validated measures from both tasks (simple and complex) with measures from structured track driving exercises with North Carolina State Highway Patrol (NCSHP) cadets. We asked whether task scores at different levels of complexity correlated with a track assessment of complex driving skills. We also used a post-test survey to compare the cadets' perceptions of the PTT to their initial experience on the track. We evaluated the perceived effectiveness of the complex PTT in a professional driver training curriculum. We also included a qualitative analysis of how the complex task was integrated with training at various EMS and police customer sites.

STUDIES

Informal Assessment of a PTT

We conducted an informal assessment of current Profiler users to determine how trainers integrate a PTT into their learning environments. The students in our sample were undergoing driver training for EMS, police driver training, bus driver training, or patrolling, and the users were distributed across the country.

In general, users found the PTT beneficial in accurately assessing reaction time and scanning skills. The PTT was identified as particularly useful to evaluate driving ability without risk to vehicles, for periodic sustainment training, for remedial training after a student performs poorly on a track, and by utilizing students' time more efficiently. Thus, the application becomes most useful when employed in conjunction with in-class training and live training (e.g., track training or commentary driving).

When there were concerns with using the PTT, they stemmed either from user expectations (e.g., believing the application would be more realistic than it is) or from a plateau in learning. When students understood the application's purpose (as an assessment tool for specific skills used while driving, rather than a video game), they became acutely aware of their driving skills and were motivated to practice to proficiency. When students were asked to run through the PTT numerous times within short time windows, they quickly became good at scoring well and somewhat bored with the application. We are finding support for this learning plateau as well. Several volunteers affiliated with Research Triangle Institute (RTI) have run the Profiler application approximately every week for at least five weeks. Though data collection is still underway, we have found that Profiler scores increase markedly between the first and second run-throughs, the scores increase somewhat between the second and third, but by the fifth run-through participants essentially reach a maximum score. Though these data are preliminary, they suggest that PTT's are useful for gaining some basic driving skills, or assessing some abilities, on an initial and periodic basis, but their limits (e.g., in realism, in capabilities) rapidly become drawbacks to students when they are used often over short periods of time.

Correlational Validation Study

The primary objective of our first study was to validate Profiler scores (and, therefore, its use as a PTT) in relation to objective measures of driving on the track. The secondary objective was to compare the basic psychometric properties of Profiler against the better-studied POL [25,26]. Our previous studies were limited by the reliance on subjective ratings, not formal evaluation, of cadets' driving behavior using scales that the instructors used for training. An objective driving course was set up with the cooperation of the NCSHP at their training center track in Garner, NC.

Participants. The participants were 50 cadets from the 100th NCSHP training academy class, ages ranging from 21–33 years with a mean age of 25.9 years. Heights ranged from 66-75 inches, with a mean of 70.6 inches, and weights from 148–215 pounds, with a mean of 180.0 pounds. Computer and track data were available for all participants. All participants read and signed an informed consent form describing the risks and benefits of participation.

Computerized Assessment. Prior to track testing, one POL and two Profiler tests were administered. A complete description of POL is given in detail elsewhere [25]. Briefly, POL is a computerized task that presents 36 randomly generated 3-second successive displays to the participant. The test has two response requirements on each display. A tap on the keyboard space bar indicates one of four combinations of "headlights" and "taillights" appearing at the center of the screen. The secondary (outer) task requires a response on one of the four keyboard arrow keys indicating the direction (and location) of a stop sign octagon that can occur in one of 12 fixed locations. The octagon can, thus, appear left, right, up, or down and at three concentric radii outward from the central display. Space bar or arrow key responses may or may not be required on each 3-second display. A divided-attention display occurs when the participant must respond to both central and outer stimuli. During testing, task difficulty increases over five levels by decreasing the discriminability of the octagon relative to the other shapes and colors in the display. The test is self-administered with on-screen instructions that insure participants respond comparably to each type of display.

A composite score from each of the five difficulty levels of POL was compared with track measures. The composite score was a linear combination of speed and accuracy of space bar and arrow key responses to central, single target, and divided-attention responses at three increasing rings from the central display.

Profiler uses the POL scoring algorithm and presents the central and outer displays as part of a filmed lap around the NCSHP driving track. The test is also self-administered, progressing through five levels of difficulty. The participant drives with an appended steering wheel and foot pedals. Similar scores to POL are gathered from each level: scanning (single target tracking at increasing distances from the center of the screen), divided-attention (multiple target tracking at increasing distances from the center of the screen), and an overall score. Figure 1 illustrates a sample output from Profiler.

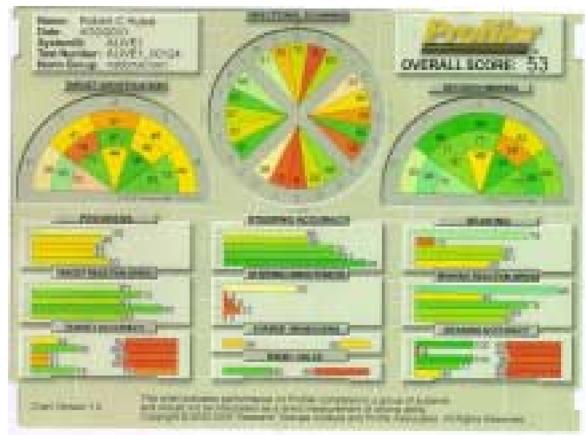

Figure 1. Sample Profiler Output

Profiler was administered a second time after track training to evaluate the cadets' perceptions of the PTT experience in comparison with their testing on the track. Ten 7-point scales assessed exercise difficulty, entertainment level, scanning skills, multi-tasking skills, and awareness of tunnel vision.

Driving Evaluation. The driving track consisted of two ovals sharing a common straightaway. The larger oval had simple, long multiple apex turns and supported higher driving speeds. An obstacle course with 11 cone "stations" was set up on both ovals for the study. Each cadet drove the track without an instructor in a Ford police cruiser equipped with a roll cage.

Three types of obstacles were used to evaluate driving skills. The first type of obstacle was a stop sign; cadets were rated on whether or not they made a complete stop. The second type of obstacle was an evasive action. Each of the six evasive action exercises had a choice of three lanes marked by four rows of cones. The visual approaches to the evasive action sections were obscured by vehicles or hills, precluding a full view of the three lanes until approximately 50 feet from the entrance. A pair of small red target cones were placed at the center and end of one of the three lanes; cadets were scored on whether or not they chose the correct lane, and the number of marker cones touched or displaced. The third type of obstacle was a lane change. The obstacles were placed around the track so that driving speed could not exceed 60 mph between any two sections. A driving instructor at each station recorded the number of cones hit or misplaced and whether or not the cadet completed the maneuver successfully. Lap times were also recorded.

Results. The mean composite scores for the cadets' performance on POL and both Profiler tests are shown in Figure 2. The correlation between the POL and Profiler mean values was 0.41 (first Profiler test) and 0.49 (second Profiler test). Profiler scores exhibited the same degradation in performance as POL as difficulty levels increased, suggesting that the presentation format does not affect psychometric properties. However, test scores on the first Profiler administration reflect a more complex and visually demanding presentation.

There was marked improvement from the first Profiler test to the second across all five levels of difficulty. The correlation between mean values on the two Profiler tests was 0.95. The test-retest reliability from the first Profiler test to the second was, for levels 1-5, respectively, 0.57, 0.61, 0.53, 0.68, and 0.52.

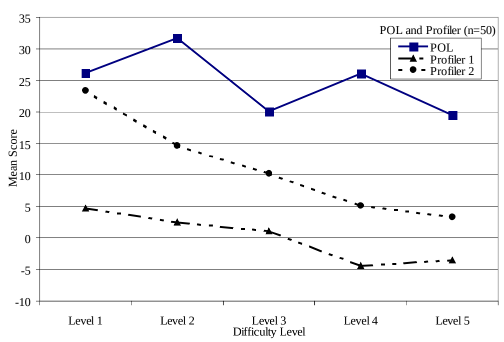

Figure 2. Cadet POL and PTT Performance

Figure 3 shows the cadets' performance during the objective driving evaluation. The first column (black) illustrates that the number of cones hit in evasive action maneuvers increased for the second lap and decreased for the third. The sum cones measure (diagonal bars) reflects a sum of all cones hit during the track evaluation, and includes evasive actions segments, lane changes, and stop signs. The sum cones measure showed a similar pattern as observed with the evasive action errors. While errors increased on the second lap, lap times (corrected, the third column) decreased marginally over the three test laps.

Table 1 shows the relationships between the computerized test scores over five levels of difficulty and the number of cones hit during the evasive action (EA) maneuvers on the test track. Generally, scores on the second Profiler administration were similar to the first, but correlation values with track performance were weaker. Note that a negative correlation between computerized scores and track performance reflects that the cadets with higher computer scores showed fewer driving errors on the track. Table 1 shows that as both the computer testing (levels 4 & 5) and the track testing (lap 2) became more difficult, the correlations increased. Most of the correlation numbers for evasive action cones and sum cones were in the same direction, with the evasive action cones showing a slightly stronger relationship to computer scores than sum cones. To simplify the table, only evasive action (EA, cones hit for each lap) correlation numbers are shown.

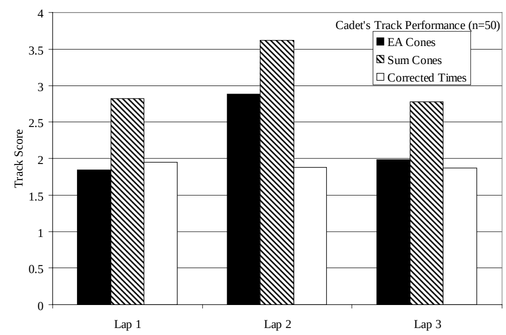

Figure 3. Cadet Driving Performance

Table 1. Relationships between Composite PTT Scores over 5 Levels of Difficulty and Track Scores

PTT/ Level	EA Lap 1	EA Lap 2	EA Lap 3
POL/1	-.12	-.31	-.30
POL/2	-.14	-.27	-.08
POL/3	-.26	-.32	-.01
POL/4	-.16	-.43	.03
POL/5	-.25	-.39	.04
Profiler/1	.11	-.17	-.06
Profiler/2	.29	-.05	.19
Profiler/3	.18	-.09	.04
Profiler/4	.03	-.27	.02
Profiler/5	-.09	-.24	.09

The track correlation values indicate that driving skills on the second lap of the objective driving course were related to both POL and Profiler scores, generally at the more difficult levels (4 & 5) of both tests, where stronger negative relationships appear between test scores and the number of cones hit.

Discussion. The study results suggest that an interactive PTT using complex, integrated responses over increasing levels of difficulty is able to assess complex perceptual skills, some of which were necessary for police cadets to negotiate complex maneuvers on a closed-circuit track. The data suggest that an interactive PTT was able to present complex patterns of visual search at procedural and strategic levels: cadets had to scan ahead, anticipate distant hazards, respond quickly, and perform more than one task at a time to improve their scores. In both simple and complex tasks, composite scores decreased linearly with level of task complexity and difficulty. Skills were practiced in the complex PTT over two 20-minute sessions; scores improved from the first administration to the second, over all levels of difficulty, suggesting a beneficial result of repetitive practice. However, this study did not formally address transfer of training issues.

Importantly, the cadets who performed better on both the simple task and the complex PTT generally showed fewer errors on the track evaluation. The relationship was most evident between the second of three laps and the complex levels of both computer assessments. During the track training, the cadets used the first lap to familiarize themselves with the track layout, and subsequently pushed themselves to improve times on the second and third laps. This self-induced pressure was enough to increase driving errors, similar to the increasing computerized task demands: on the second lap, a small improvement in lap times was accompanied by a substantial increase in driving errors. As the lap times did not improve noticeably from the second to third lap, and errors decreased on the third lap, much of the cadets' overall skill acquisition occurred during the second lap.

The modest correlation values between computer and track performance suggest that similar, but not identical, skills were assessed in both exercises. Interestingly, the "objective" evaluation of track skills (a relatively costly and time consuming process) did not produce better relationships between lab assessment data and scores generated by instructors using paper and pencil rating scales [26]. Since the instructor rating scales were far less costly to use, our future studies will focus on refining instructor-rating instruments rather than setting up complex evaluation procedures for driving skills.

The post-test comparison of perceptions of the computer and track exercises showed no significant differences for any ratings of level of difficulty, entertainment, learning to scan, learning to do multiple tasks, and learning to reduce tunnel vision. The lack of differences could indicate a bias in that the scales only assessed perceptions about the specific skills addressed in the PTT. The data suggest, however, that this skill set can be effectively presented from the cadet's point of view in an interactive video production. Apparently, the more unfamiliar experience that a computerized test presents, the closer it approximates the experience of driver training under pressure. The first exposure to the complex PTT, while producing lower scores, showed a stronger relationship to track errors. This suggests (as do the indications of a plateau in learning cited in our informal assessment) that as the cadets became familiar with the task requirements and less surprised by the demands of the production, the scores were less likely to reflect "live" training experiences.

Replication of Correlational Study

In a replication and extension of the correlational validation study, fifty cadets from the 104th NCSHP training academy participated concurrently with their driver training on the track. In addition to the pre- and post-test instruments, NCSHP driving instructors also rated the cadet driving skills with 18-item seven-point scales on four track exercises. The skills for each track exercise included items such as steering, lane control, quality of turns, hazard recognition, ability to listen to instructions, and smoothness of vehicle operation. The four exercises were advanced pursuit (AP), primary/secondary pursuit day (PD), primary secondary pursuit night (PN), and a final exercise of advanced primary/secondary pursuit (APS). Forty-six of the cadets self-administered the Profiler test two times, while 45 completed both Profiler tests and all track evaluations.

We were interested in whether or not we could find a relationship between PTT scores and instructor track ratings. If we found a relationship, we were interested in which category or categories of rating across track exercises were related to computerized testing. We also looked carefully at the administration of two Profiler tests. We developed two somewhat competing hypotheses, either that the difference between the two Profiler scores is informative in assessing driver skills, or that the Test 2 score itself is most informative (the Test 1 serving mainly as a means for the student to become comfortable with the assessment environment).

The test-retest correlation between computerized assessments was 0.81. The scores showed a similar degradation over difficulty levels as reported in Figure 2. On average, the Test 1 scores were 47% of Test 2 scores, suggesting a fairly steep learning curve. The correlations between PTT tests and track scores were weak, ranging from 0.03-0.25, all positive and not significantly different from zero. Overall, the Profiler composite scores from Test 2 were more strongly related to track scores (r = 0.25 AP; 0.05 PD; 0.16 PN; 0.15 APS), again all positive but not significantly different from zero. The data tend to support our hypothesis that the second Profiler test gathered information that was closely linked with a live training exercise. This suggests that as performance was more focused on visual and cognitive skills, rather than "game-learning" skills, a stronger relationship to track performance emerged.

The difference between Test 1 and Test 2 scores was not significantly related to any of the track measures. While the correlations between the difference and the track scores (r = 0.01 to 0.15) were positive, none was significantly different from zero. This indicates that learning to improve on the computerized task was independent of live training skills.

Test 1 and Test 2 Profiler scores were normally distributed. We identified "upper" and "lower" performers based on a ±1 standard deviation cutoff from the mean of Test 2, and found that the relationship between computer scores and track scores emerged more strongly. Figure 4 shows the mean track ratings for the four track exercises. The upper and lower differences for AP, PD, PN, and APS are not significantly different from one another but all are in the expected direction.

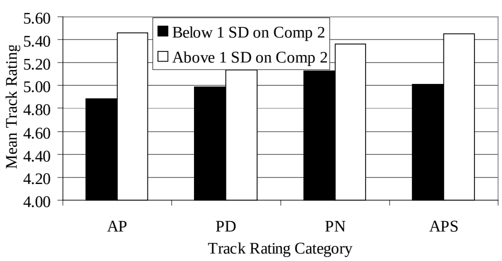

Figure 4. Track Scores for Test 2 Scores 1 Standard Deviation Above and Below the Mean

For most of the specific skills rated by track instructors (16 out of 18), mean differences were not significant, implying that Profiler does not consistently measure those skills. This might be expected; included in the list are driving posture, minimizing recklessness, and radio use. However, the instructors mean rating across all four track exercises for smoothness of vehicle operation was significantly different for the upper and lower performance on Test 2 (F=10.11, p<0.006), and a similar difference approached significance for setting up for turns (F=4.24, p<0.06). Further data will enable us to explore these comparisons to understand the complex relationship between PTT testing and driving. It appears, though, a limited set of skills that reflect a driver's skill in being smooth and calm under high demand was moderately related to scores on a computer task that presented similar demands. The findings suggest that a cadet's ability to stay calm may be a common factor in his or her ability to do well in high demand simulated or live environments.

The results from both POL and Profiler testing suggest that pre-testing of driving skills in a controlled environment may have some utility in assessing and predicting real-world driving skills. We viewed these results as preliminary and began further studies of integrating the computerized tests into a driving curriculum.

Integration Study

We are currently conducting a formal study with emergency response driver trainers of how instructors can and should integrate the PTT into their curricula. In North Carolina, law enforcement driver training (both pre-service and ongoing in-service) is handled by area community colleges. In addition to the NCSHP participants from the previous studies, we recruited driver trainers and students from Asheville-Buncombe Technical Community College in Asheville, NC, and Western Piedmont Community College in Morganton, NC. We also recruited trainers specializing in other forms of emergency response, such as fire and EMS. Finally, we recruited driver trainers from the North Carolina Justice Academy in Salemburg, NC.

We present each student with a questionnaire before using the PTT application, and after. The pre-test instrument addresses topics such as familiarity with computing applications and video games, driving history, and expectations of the PTT. The post-test instrument gathers ratings of the application. We also developed a questionnaire for instructors to complete in addition to a structured interview by experimental observers. Preliminary results described next are from the NCSHP cadets and from six Asheville fire fighters (average age 37.2 years, average driving experience 21.7 years); we are currently setting up procedures to collect additional student data.

The majority of NCSHP study participants (35/50 cadets) answered that, through the PTT, they became more aware of their reaction and scanning skills. Few were "surprised" by Profiler, and when asked for means of improvement nearly all focused on making the application more realistic (e.g., improved steering, use of the accelerator pedal, adding night driving, more radio traffic and ambient sounds).

All fire fighter participants reported that the application measured the visual and mental demands of emergency driving. On a 5-point scale (1-Not at all, 5-Very much) they felt that the PTT somewhat to moderately met their expectations (see Table 2). The participants rated the simulated driving experience as somewhat to moderately challenging (3.7/5), and they felt the PTT represents a meaningful addition to training (3.8/5) and ongoing practice (3.3/5).

Table 2. Ratings of PTT Meeting Expectations

Category	Mean
Improving decision making	3.3/5
Performing more than one task at a time	3.5/5
Improving visual scanning	3.8/5
Staying calm w/ increased distraction	3.3/5
Improving steering ability	2.5/5
Improving ability to execute evasive maneuvers	2.5/5
Improving reaction time	3.3/5
Reducing tendency to tunnel	3.3/5

Analyses from the NCSHP cadets suggest that none of age, education, driving experience, computer usage, self-ratings of proficiency with computers, or video gaming experience consistently affect PTT scores (all $F<1.7$). We infer from these results that the computerized assessment is equally applicable and relevant to all students.

RIGHT MIX OF CONSTRUCTIVE AND LIVE ENVIRONMENTS

Training Triangle and FAPV Model

We have developed a "Training Triangle", shown in Figure 5, that shows a relationship between level of proficiency and relative amount of time spent in didactic, virtual, constructive, and live learning environments [17]. We use the terms familiarization, acquisition, practice, and validation to describe level of proficiency. Becoming familiarized with to-be-learned material implies acquiring knowledge about components (e.g., machinery and equipment for a diagnostic trainer, or the rules of the road for a driver trainer), their capabilities, and their characteristics (e.g., location, expense, importance, length, value to society). Familiarization is relatively passive; knowledge can be gained, for instance, by absorbing a presentation or through reading. Acquiring a skill is learning techniques and procedures. Students must actually perform the skills to truly acquire them. However, acquisition can normally be achieved in a virtual environment. Practice is an extension of acquisition. During practice (often called proceduralization), the student internalizes techniques and procedures by performing the skill. Practice can begin in a virtual environment, extended practice can occur in a constructive environment, and focused practice can occur in a live environment, when available. Skills are validated when students are tested on their ability to perform the skills. The more realistic the setting in which skills are validated, the more confidence both instructor and student will have in their applicability on the job.

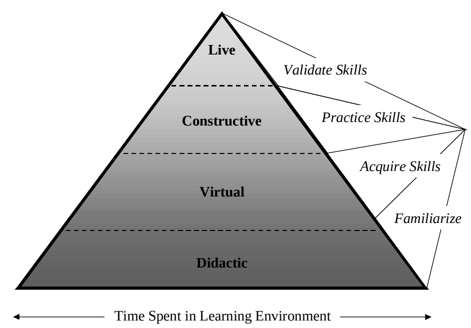

Figure 5. Training Triangle

Driver Training

Though our studies are ongoing, we are gaining an awareness of how and when a PTT that assesses key scanning and reaction skills is useful during driver training. Because desirable characteristics of a PTT for perceptual training with inexperienced drivers include task repetition, gradually increasing demands, and unpredictable scenarios, the PTT is best employed twice or perhaps three times for most students; beyond that the scenarios become predictable and task demands lessen. The PTT also should be used after the student has gained some driving experience (that is, after the student has gained familiarity with driving), so that its applicability and realism is enhanced, but before intensive specialized training (such as emergency or high-speed driving on a track), since the PTT highlights the student's scanning and reaction tendencies, which can be used to focus on-track training.

The PTT is a constructive device, hence is useful when actual equipment (in this case, vehicles) is expensive and/or difficult to obtain. It is but one tool, though, in the driver trainer's toolkit; it is most effective when used in conjunction with classroom and on-track time. In sum, we envision the PTT as most useful for practice rather than acquisition of skills, with its results feeding into validation of complex perceptual driving skills.

ACKNOWLEDGEMENTS

This research was supported by grant AA10180 from the National Institutes of Alcohol Abuse and Alcoholism (NIAAA). The research was also supported by the Training Academy of the NCSHP. The opinions and conclusions in this paper do not reflect the views of the sponsoring agencies.

REFERENCES

1. Amato, I. (2001). Helping Doctors Feel Better. Technology Review, 104(3), 64-48.
2. Ball, K., Beard, B., Roenker, D., Miller, R., & Griggs, D. (1988). Age and visual search: Expanding the useful field of view. Journal of the Optical Society of America, 5, 2210-2219.
3. Ball, K., & Owsley, C. (1991). Identifying correlates of accident involvement in older drivers. Human Factors, 33, 583-595.
4. Barnett, B., Helbing, K., Hancock, G., Heininger, R., & Perrin, B. (2000). An Evaluation of the Training Effectiveness of Virtual Environments. Presented at the Interservice/Industry Training, Simulation and Education Conference. November 30, 2000, Orlando, FL.
5. Bédard, J., & Chi, M.T.H. (1992). Expertise. Current Directions in Psychological Science, 1(4), 135-139.
6. Brown, F.J. (2000). Preparation of Leaders. Institute for Defense Analyses Report, IDA Document D-2382, January, 2000.
7. Camburn, D.P., Gunther-Mohr, C., & Lessler, J.T., (1999). Developing New Models of Interviewer Training. International Conference on Survey Nonresponse, Portland, OR, October 28-31, 1999.
8. Chapman, P.R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. Perception, 27, 951-964.
9. Christianson, S. (1992). Emotional stress and eyewitness memory: A critical review. Psychological Bulletin, 112(2), 284-309.
10. Crundall, D.E, Underwood, G., & Chapman, P.R. (1998). How much do novice drivers see? The effect of demand on visual search strategies in novice and experienced drivers. Underwood, G. (Ed.) Eye Guidance in Reading and Scene Perception, Elsevier Science Ltd., New York. 395-417.
11. Decina, L.E., Gish, K.W., Staplin, L., & Kirchner, A.H. (1996). Feasibility of new simulation technology to train novice drivers. NHTSA, DOT, DOT HS 808548.
12. Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66(3), 183-201.
13. Ericsson, K.A., Krampe, R.T., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review, 100(3), 363-406.
14. Frank, G.A., Helms, R. & Voor, D. (2000). Determining the Right Mix of Live, Virtual, and Constructive Training, Proceedings of the 21st Interservice/Industry Training Systems and Education Conference, Orlando, FL.
15. Helms, R.F., Hubal, R.C., & Triplett, S.E. (1997). Evaluation of the Conduct of Individual Maintenance Training in Live, Virtual, and Constructive (LVC) Training Environments and their Effectiveness in a Single Program of Instruction. Final Report, September 30, 1997. Submitted to Battelle RTP Office, Subcontract # TCN 97031, Delivery Order #0027, Dated April 16, 1997.
16. Hill, R., Chen, J., Gratch, J., Rosenbloom, P., & Tambe, M. (1998). Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft, Proceedings of the Seventh Conference on Computer Generated Forces & Behavioral Representation, May 12-14, 1998, Orlando, FL.
17. Hubal, R.C., Frank, G.A., & Guinn, C.I. (2000). AVATALK Virtual Humans for Training with Computer Generated Forces. Proceedings of the Ninth Conference on Computer Generated Forces. Institute for Simulation & Training: May 16-18, 2000, Orlando, FL.
18. Hubal, R.C., & Frank, G.A. (2001). Interactive Training Applications using Responsive Virtual Human Technology. To be presented at the Interservice/Industry Training, Simulation and Education Conference. November 26-29, 2001, Orlando, FL.
19. Lerner, N.D, Tornow, C.E., Freedman, M., Llaneras, R.E., Rabinovich, B.A., & Steinberg, G.V. (2001). Preliminary investigations of highway design countermeasures to aid drivers with limited experience: Final Report. Contract DTFH61-98-C-000063. Federal Highway Administration, U.S. Department of Transportation.
20. Loftin, R.B., & Kenney, P.J. (1994). Virtual Environments in Training: NASA's Hubble

Space Telescope Mission. 16th Interservice/Industry Training Systems & Education Conference, Orlando, FL.
21. Maltz, M., & Shinar, D. (1999). Eye movements of younger and older drivers. Human Factors, 41(1), 15-25.
22. Massie, D.L., Campbell, K.L., & Williams, A.F. (1995) Traffic accident involvement rates by driver age and gender. Accident Analysis and Prevention, 21(1), 73-87.
23. Mayhew, D.R., & Simpson, H.M. (1995). The role of experience: Implications for the training and licensing of new drivers. Traffic Injury Research Foundation of Canada.
24. Mills, K.C. (2000). Tunnel Vision: Lifeline or Killer? Training to Keep Your Cool. Law and Order, November, 2000.
25. Mills, K.C., Parkman, K.M., & Spruill, S.E. (1996). A PC-based software test for measuring alcohol and drug effects in human participants. Alcoholism: Clinical and Experimental Research, 20(9), 1582-1591.
26. Mills, K.C., Parkman, K.M., Smith, G.A., & Rosendahl, F. (1999). Prediction of driving performance through computerized testing: High-risk driver assessment and training. Transportation Research Record, 1689, 18-24.
27. Mills, K.C., Spruill, S.E., Kanne, R.W., Parkman, K.W., & Zhang, Y. (in press). The influence of stimulants, sedatives and fatigue on tunnel vision: Risk factors for piloting and driving. Human Factors.
28. Mourant, R.R., & Rockwell, T.H. (1970). Mapping eye-movement Patterns to the visual scene in driving: An Exploratory study. Human Factors, 21(1), 81-87.
29. Mourant, R.R., & Rockwell, T.H. (1972). Strategies of visual search by novice and experienced drivers. Human Factors, 14(4), 325-335.
30. Perryman, K.M., & Fitten, L.J. (1996). Effects of normal aging on the performance of motor-vehicle operational skills. Journal of Geriatric Psychiatry and Neurology, 9, 136-141.
31. Ranney, T.A. (1994). Models of driving behavior: A review of their evolution. Accident Analysis and Prevention, 26(6), 733-750.
32. Shanteau, J. (1992). How Much Information Does an Expert Use? Is It Relevant?. Acta Psychologica, 81, 75-86.
33. Szlyk, J.P., Seiple, W., & Viana, M. (1995). Relative effects of age and compromised vision on driving performance. Human Factors, 37(2), 430-436.