Technical Report 1331

Transferring from the Simulator to a Live Robotic Environment: The Effectiveness of Part-task and Whole-task Training

Patricia L. McDermott
Thomas F. Carolan
Alia Fisher
Mark R. Gronowski
Andrew M. Gacy
Alion Science and Technology, Inc.

October 2013

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

Alion Science and Technology, Inc.

Technical review by

Lynne Rochette, U.S. Army Research Institute
James Belanich, Institute for Defense Analyses

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZXM, 6000 6th Street (Bldg 1464 / Mail Stop: 5610), Fort Belvoir, Virginia 22060.

FINAL DISPOSITION: Destroy this Technical Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy): October 2013	2. REPORT TYPE: Interim	3. DATES COVERED (from. . . to) September 2009 to September 2013

4. TITLE AND SUBTITLE Transferring From the Simulator to a Live Robotic Environment: The Effectiveness of Part-task and Whole-task Training	5a. CONTRACT OR GRANT NUMBER W91WAW-07-C-0081
	5b. PROGRAM ELEMENT NUMBER 611102
6. AUTHOR(S) Patricia L. McDermott, Thomas F. Carolan, Alia Fisher, Mark R. Gronowski, and Andrew M. Gacy	5c. PROJECT NUMBER B74F
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER 2902
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Alion Science and Technology, Inc. 4949 Pearl East Circle Boulder, CO 80301	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1331

12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A;
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter POC: Karin A. Orvis

14. ABSTRACT *(Maximum 200 words)*:

Part-task training methods are widely used in training when the full target task is considered too complex or impractical for training as a whole task. However, part-task training has had mixed success in transfer to the whole task. Limited research has focused on part tasks that do not interact but are performed concurrently in the whole task. The present research was designed to help address this gap in the context of an Army-relevant task and training situation: teleoperating a robotic device to detect and identify vehicles. The experiment involved training in a simulated environment and transfer to performance with a live robotic system. Training methods included part- and whole-task training in the simulation environment and part-task training in the live environment. Results indicated a significant benefit in vehicle mobility skills for the condition receiving part-task training both in the simulated and live environment. Participants in this training condition also reached transfer performance criteria in fewer scenarios, and the benefit to mobility persisted in the final "freeform scenario." Additionally, part-task training in the live environment, prior to the transfer scenarios, was more efficient than other conditions, resulting in less time spent training on the live robotic system.

15. SUBJECT TERMS
Transfer of Training, Robotics Training, Training Effectiveness, Simulation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	46	Dorothy E. Young (703) 545-2316

Technical Report 1331

Transferring from the Simulator to a Live Robotic Environment: The Effectiveness of Part-task and Whole-task Training

Patricia L. McDermott
Thomas F. Carolan
Alia Fisher
Mark R. Gronowski
Andrew M. Gacy
Alion Science and Technology, Inc.

Foundational Science Research Unit
Gerald F. Goodwin, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg. 1464
Fort Belvoir, VA 22060

October 2013

TRANSFERRING FROM THE SIMULATOR TO A LIVE ROBOTIC ENVIRONMENT: THE EFFECTIVENESS OF PART-TASK AND WHOLE-TASK TRAINING

EXECUTIVE SUMMARY

Research Requirement:

This experiment was conducted as part of an ongoing United States Army Research Institute for the Behavioral and Social Sciences (ARI) research project to identify evidence-based guidelines for the relative effectiveness of different training methods for acquiring and transferring cognitive skills involved in complex task domains. This experiment's emphasis was on strategies for managing learner workload during training while effectively promoting transfer of learned skills.

Part-task training involves methods for reducing the complexity of a task during training with the objective of improving the effectiveness and/or efficiency of skill learning and transfer to the whole task. Part-task methods are widely used in training programs when the full target task is considered too complex or impractical to start training on initially. Often the objectives of part-task training include reducing training time and costs without negatively impacting learning and transfer to the whole task in an operational environment. However, part-task training has had mixed success in transfer to the whole task, in comparison to whole-task training. Much of the part-task training research has focused on part tasks that are segments of a serial task or components that are integrated and performed concurrently in the whole task. There is limited research focused on part tasks that are separate but concurrent in the whole task. This domain includes a variety of task types that involve vehicle control and navigation while monitoring the environment to detect and identify potential threats. The present research was designed to help address this gap in the context of an Army-relevant task and training situation. The task is teleoperating a robotic device to detect and identify vehicles as friendly, enemy, or neutral. The training involved a simulated environment, where only part-task training may be available, and transfer to performance with a live robotic system and test environment.

Procedure:

Thirty-nine participants, 24 males and 15 females, learned to remotely drive a small unmanned ground vehicle while looking for vehicles and identifying those vehicles as friendly, enemy, or neutral. The task consisted of three component parts: mobility, target detection, and target identification. Participants were trained using a simulated vehicle and environment, which then transitioned to a live ground vehicle in a live environment. Four different training conditions were used in the simulation training environment: (1) part-task, (2) whole-task, (3) part-task followed by whole-task, and (4) a fourth condition involved part-task training in the simulation training environment as well as part-task training in the live environment prior to starting transfer scenarios in the live environment.

Transfer was to the whole task in the live environment one week later – measured using two types of transfer scenarios. The first transfer scenario was a traditional transfer of learned skill paradigm, but also assessed one's ability to quickly get to a proficient level of performance on a new system. After this first transfer scenario was completed, all participants performed a second transfer task (i.e., the “freeform scenario”), in which the objective was to identify all threats in the environment, but without a fixed route through the environment. Key transfer performance criteria included: 100% detection and identification accuracy, with no collisions or wrong turns, and within a specified time limit.

Findings:

Twenty-nine participants achieved transfer performance criteria in the live environment within the allocated timeframe. The four training conditions were equally effective at getting 70-80% of the learners to criterion level performance. Results from the first transfer scenario indicated a significant benefit to mobility skills (i.e., fewer collisions) for the condition receiving part-task training in both the training simulation and live environment. Participants in this condition also reached transfer performance criteria in significantly fewer scenarios, and the significant benefit to mobility persisted in the freeform scenario. Examining the scenario performance plots suggests that this benefit was not due solely to the extra training time but also to the part-task approach. In addition, completing part-task training in the live environment (prior to the transfer scenarios) was more efficient than other conditions in terms of training time on the live robotic system. The condition with live part-task training spent an average of 46% less time on the live robotic system than participants in the other three conditions who started directly with a whole-task transfer scenario.

Utilization and Dissemination of Findings:

The results of this experiment provide information for training researchers and training designers. The results suggest that part-task training in the live environment can increase the effectiveness and efficiency of learning time/attention sharing skills in the whole task, when subtasks are concurrent but separate tasks. For the military trainer, the results add to the research base that suggests that part-task training can be used to reduce training cost on operational equipment without sacrificing transfer performance or efficiency.

TRANSFERRING FROM THE SIMULATOR TO A LIVE ROBOTIC ENVIRONMENT: THE EFFECTIVENESS OF PART-TASK AND WHOLE-TASK TRAINING

CONTENTS

Page

FIGURES

TABLES

TRANSFERRING FROM THE SIMULATOR TO A LIVE ROBOTIC ENVIRONMENT: THE EFFECTIVENESS OF PART-TASK AND WHOLE-TASK TRAINING

Introduction

The present research was conducted as part of an ongoing project to identify evidence-based guidelines for the relative effectiveness and efficiency of different training methods for acquiring, retaining, and transferring cognitive skills involved in complex Army-relevant task domains (Carolan, Belanich, McDermott, Hutchins, & Wickens, 2011). For a given task, what training methods or combinations of methods are more effective in terms of performance, retention, and especially transfer outcomes and/or reduction in training time to achieve criterion outcome levels?

Literature review and meta-analyses were conducted to systematically evaluate the effectiveness of training methods in the context of key moderating factors, including transfer criteria, task factors, and individual differences in ability or experience. The results from the review and meta-analyses also identified areas where the research evidence is limited or ambiguous. These gaps informed subsequent experiments designed to address training method-specific research questions as applied to basic and complex cognitive skills.

Research on methods for training complex cognitive tasks emphasizes the role of learner effort and workload in training and learning (e.g., Paas & van Gog, 2009). While training tasks that are too difficult for the experience level of the learner can inhibit learning, there is evidence that training tasks that require more learner effort may yield higher performance during transfer, especially transfer to new problem situations. Accordingly, a key research emphasis for the experiments conducted as part of this ongoing project was on strategies for effectively managing learner effort and workload during training and on research comparing the effectiveness of training methods for promoting transfer performance.

Part-Task Training

Part-task training involves methods for reducing the full complexity of a target task during training with the objective of improving the effectiveness and/or efficiency of learning and transfer to the whole target task. Part-task methods are widely accepted and used in training programs when the full target task is considered too complex or impractical to start training on initially (Lintern, 1991). Part-task training is often used to reduce training costs by replacing some of the training on expensive operational equipment or full mission simulators with training using low cost training media. Another argument for part-task training is that it can save whole-task learning time, speeding up the learning process and increasing training efficiency relative to whole-task training (Wightman & Lintern, 1985).

Part-task training has a long research history primarily focused on complex manual control and tracking tasks. Wightman and Lintern (1985) reviewed the literature on part-task training of psychomotor tasks and identified three strategies for reducing the complexity of a target task: segmentation, fractionation, and simplification. Segmentation involves separating component tasks that are performed sequentially in the whole task and training the sequential part tasks, such as the sequential steps in a surgical procedure. Fractionation involves separating tasks that are performed concurrently in the whole task and training the fractional part tasks, such as controlling heading and altitude in flight control or the gear shift and clutch in manual shifting. Simplification, as the name implies, involves keeping all the parts but simplifying the full task complexity by reducing fidelity, time pressure, number of targets, or providing worked examples or other scaffolding techniques.

In their review, Wightman and Lintern (1985) conclude that part-task training has received limited empirical support. In comparison to whole-task training, part-task training has had mixed success in transfer to performance on the whole task. Part-task training provided a relative transfer benefit when segmented part tasks were sequential in the whole task, but not when fractionation part tasks were performed concurrently in the whole task. The lack of effectiveness of part-task training for concurrent part tasks is thought to be due to the need to practice the timesharing between the concurrent tasks that can only be practiced in the whole task (Lintern & Wickens, 1991). Therefore, transfer to the whole task requires additional time to develop timesharing skills and integration of part tasks into the whole task (Goettl & Shute, 1996).

A series of experiments using a complex task (Space Fortress computer game) has provided some support for the value of part-task training with complex cognitive tasks (Lintern, 1991). The task involves concurrent and coordinated use of perceptual-motor skills with conceptual, procedural, and strategic knowledge components (Frederiksen & White, 1989). Training on task components, when it involved additional training prior to whole-task training (e.g., Mané, Adams, & Donchin, 1989), resulted in better performance throughout the whole-task practice. Analysis showed no significant differences in training times (Lintern, 1991). Part-task training strategies that incorporated opportunities to practice concurrent part-task integration and timesharing were also effective relative to whole-task training (e.g., Gopher, Weil, & Siegel, 1989).

A recent meta-analysis of the part-task training research (Wickens, Hutchins, Carolan & Cumming, 2013), also conducted as part of the aforementioned ongoing project, included 22 experiments that provided statistical information necessary for effect size analysis. The tasks in these experiments involved procedural, perceptual-motor, and cognitive components and included tasks such as multi-dimensional tracking, simulated vehicle control with concurrent monitoring and responding to stimuli, as well as Morse code, poetry, and preparing legal presentations. The meta-analysis investigated the effectiveness of part-task training methods, relative to whole-task training, as measured on transfer performance. These were primarily measures of performance accuracy. The results indicated that, on average, over different part-task methods, measures, and task factors, part-task training was less effective for transfer performance than training on the whole transfer task. This is consistent with previous reviews. For part tasks that are performed concurrently in the whole task, therefore requiring timesharing of task elements, training the part tasks separately resulted in significant performance decrements on the transfer task, in comparison to training the whole task. For segmented part tasks that are sequential in nature, there was no overall performance advantage or disadvantage for part-task training relative to training on the whole task. Part-task methods that provided an opportunity to develop timesharing skills were more effective relative to whole-task training. Pure part-task training, where the parts are not practiced together during training, resulted in a large negative effect relative to whole-task training for fractionated tasks but was not significantly different from whole-task training for segmented tasks. In other words, if the subtasks are sequential, part-task and whole-task training methods are, on average, equally effective for transfer performance on the whole task. However, if the subtasks are performed concurrently in the whole task, whole-task training is generally more effective than part-task training.

Research Questions

Results from the literature review and meta-analyses from the larger ongoing project suggested a number of areas where additional evidence is needed to support training recommendations related to the effectiveness of part-task training strategies. Most of the part-task research has focused on part tasks that are segments of a serial task or fractions of an integrated concurrent task, where integrated implies a physical interaction or coordination between the two concurrent subtasks, such as in flight control or shifting gears. There has been a relatively limited research focus on part tasks that are concurrent but separate in the whole task; that is, the subtasks interact only through their concurrent demands on operator attention. This domain includes a variety of task types that involve vehicle control and navigation while monitoring the environment to detect and identify potential threats.

The objective of training is to transfer what is learned and apply it in an operational environment. For most of the experiments in the meta-analysis, in keeping with the part-task paradigm, the transfer task is identical to the training task, whether trained as a whole task or using a part-task method. Only one experiment in the meta-analysis included a far transfer task (defined as a task that is different than the trained task, but requires the same set of trained skills). There were no experiments in which part-task training was done via simulation and transfer was conducted via whole-task training in a higher fidelity environment. A common objective of part-task training is to reduce training time and costs on expensive operational equipment. An alternative or additional part-task training goal may be to provide training when whole-task training is not available or practical. Given these objectives, the current experiment focused on the use of part-task training as a means to reduce training time on the target task without negatively impacting transfer performance criteria.

We developed an experiment to address research gaps identified in the meta-analyses. The experimental task involved controlling an unmanned ground vehicle to detect and identify targets in an urban environment. The task consisted of three parts: teleoperation, target detection, and target identification; requiring a combination of psychomotor, spatial, perceptual, and timesharing skills. Detection and identification are sequential tasks. Mobility and detection/identification are concurrent but not integrated tasks; that is, there is no physical interaction or coordination required—the subtasks interact only through their concurrent demands on operator attention.

The research questions assume a situation in which a real robotic system and live testing environment are not available for training in the schoolhouse and the trainer may not have access to a single simulation system that performs all components of the task. Part-task training is therefore conducted using the pure part-task method. The experiment investigates the relative effectiveness of part-task (p) and whole-task (w) training in a simulation environment, where the transfer task is performance on a real robotic system in a live test environment. Training conducted via simulation is abbreviated with a lower case p or w. Part-task training in the live environment is abbreviated with a capital P and the whole task transfer scenarios are abbreviated with a capital W. To control the scope of the experiment, not all combinations of part-task and whole-task training were investigated. Instead, conditions were chosen which allowed us to address the following research questions:

1. Given only part-task training in the simulation environment, is there a benefit to providing part-task training in the live environment prior to the whole-task training*? In other words, is there a benefit to p-PW compared to p-W?* What is the impact in terms of both transfer performance and training time on the live system?
2. If whole-task training is available on the simulation system, how does whole-task training compare to part-task training on the simulation system, in terms of transfer performance on the live system and savings of time required to reach performance criteria on the live system? *In other words is w-W more beneficial than p-W?*
3. If whole-task training is available on the simulation system, does prior part-task training on the simulation system improve performance in terms of transfer performance on the live system and savings of time required to reach performance criteria on the live system? *In other words, is pw-W more beneficial that w-W?*
4. A primary concern with pure part-task training, where the part tasks are performed concurrently in the whole task, is that the timesharing skills are not learned during the part-task training. Are there significant differences in the time required to learn the timesharing skills in the live environment for the four different combinations of part and whole-task training (i.e., *p-PW, p-W, w-W, pw-W)*?
5. Do the part and whole-task training result in any differences in performance on a second far transfer task that requires some ability to adapt learned skills?
6. Do individual differences in abilities (spatial skills), related experience (videogame experience), or demographic variables (age or gender) impact training or transfer performance or interact with training method?

Method

Experimental Task

Participants learned to remotely drive a small unmanned ground vehicle while looking for other vehicles and identifying those other vehicles as friendly, enemy, or neutral. The task consisted of three parts:

1. Mobility: teleoperation of unmanned vehicle;
2. Detection: noticing targets of interest in robot's visual field; and
3. Identification: determining whether the target vehicle was friendly, enemy, or neutral, based on cues such as color, size, gun barrel length, and whether the vehicle was tracked or wheeled.

Several military systems contain this capability or a subset of these capabilities. One operator may be assigned to perform all three tasks, particularly for a single unmanned system. At other times, the different tasks may be parsed out to different operators. One operator may teleoperate an unmanned vehicle and send snapshots of suspicious vehicles to another operator for identification. Or, one operator may monitor video feed from an autonomous unmanned vehicle and identify suspicious targets and a second operator could make the final identification. It is possible that a trainee learns these tasks on a single system or on multiple systems. In this experiment, participants learned in a simulation environment and transferred those skills to a live robotic environment in which the robot was driven through urban settings.

Experimental Design

Training was a between-subjects variable with four conditions. The conditions specified whether the simulation and the live training consisted of part-task training, whole-task training, or both. Not all combinations were used. The conditions were designed to address the specific research questions detailed above.

The first condition is part-task training with a simulation, part-task training in the live environment, and then whole-task transfer in the live environment (p-PW; see Figure 1); this simulates a situation where an operator has been trained separately on the three tasks, perhaps on three different systems, and the training approach is to train the part-task skills in the live environment before undertaking transfer to the whole task.

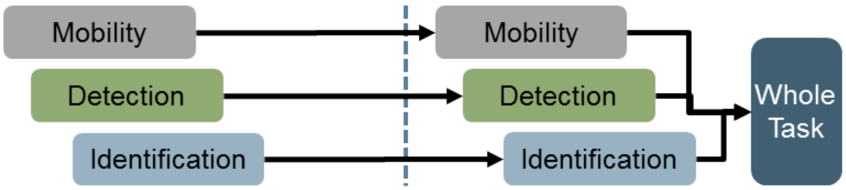

Figure 1. **The p-PW training condition: Part-task simulation training, part-task live training, and whole-task transfer in the live environment.**

The second condition, part-task training with a simulation, whole-task transfer in the live environment (p-W; see Figure 2), also simulates a situation where an operator has been trained separately on the three tasks, perhaps on three different systems. However, the training approach assumes that the part-task skills are similar enough that they will transfer to the live environment and the operator can start performing the whole task in the live environment.

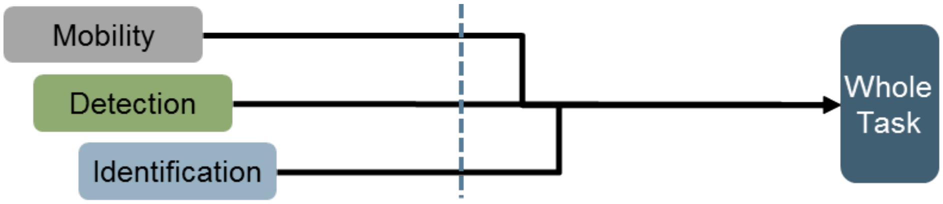

***Figure 2.* The p-W training condition: Part-task on simulation, whole-task transfer in the live environment.**

The third condition, part-task and whole-task training in the simulation with whole-task transfer in the live environment (pw-W; see Figure 3), simulates a situation where the initial training approach involved mastering the separate part-task skills individually, using a part-task training schedule, before attempting the whole-task in the simulation training system, and then transferring from the whole task in the simulation to the whole task in the live environment.

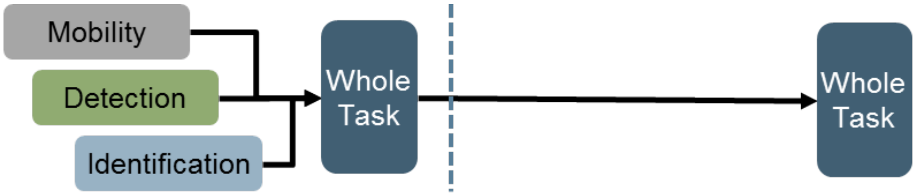

Figure 3. The pw-W training condition: Part-task and whole-task simulation training, whole-task transfer in the live environment.

The fourth condition, whole-task training in the simulation to whole-task transfer in the live environment (w-W; see Figure 4), simulates a situation where an operator is trained only the whole system in the simulation and transfer is to the whole system in the live environment.

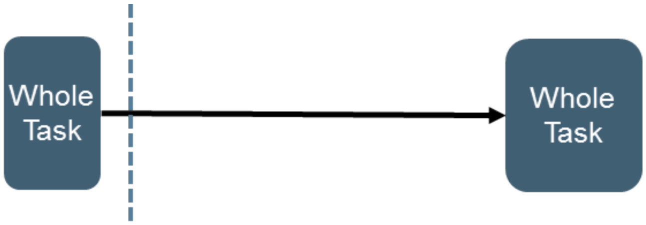

***Figure 4.* The w-W training condition: Whole-task on simulation training, whole-task transfer in the live environment.**

Comparison of conditions one and two addresses the first research question, the value of additional part-task training in the live environment, prior to whole task transfer in the live environment. Comparison of conditions two and four addresses the second research question, the relative advantages of part-task versus whole-task simulation training for transfer to the live environment. Comparison of conditions three and four addresses the third research question, the traditional evaluation of part-task training relative to whole-task training for training the whole task and for transfer to the whole task in the live environment.

All four training conditions were compared to address the fourth research question about the transfer of timesharing skills to the live environment. The transfer to the whole task in the live environment is considered a far transfer test because a different system was used (the real robotic system as opposed to the simulation used in training). In addition, there was a second far transfer task designed to evaluate flexibility or adaptability of the learned skills. In this scenario, called the "freeform" scenario, participants were not given a fixed route but were instructed to search the area to identify all enemy vehicles. The freeform scenario was the final scenario for all participants and was used to evaluate research question five. The sixth research question, regarding the interaction with individual differences, involved all four conditions.

Participants

Thirty-nine volunteers completed the experiment, 24 males and 15 females. The experiment was originally designed to have 10 participants per condition. However, upon analysis of the data from Session 1, it was realized that one participant did not pass criteria and therefore was excluded from the analysis. This resulted in only 9 participants in the w-W condition. Participants ranged in age from 18 to 39 years (*M* = 27.13, *SD* = 5.57). Participants were recruited from craigslist and from an email list of previous participants in other experiments. They were compensated $100 for their time in the 4-hour experiment.

Apparatus

The Soldier Machine Interface (SMI) was used as the primary experimental testbed. A separate software application was developed to present and log performance for part-task training. For the *detection* task, the software played a video from autonomous unmanned vehicles and participants pressed a button when they saw a vehicle. For the *identification* task, the software presented a static image and participants pressed a button to classify it as friendly, enemy, or neutral. The system logged response time, responses, and whether identifications were correct. Participants used a Logitech Dual Action game controller and a mouse to interact with the systems. Teleoperation (i.e., the *mobility* task) was done on an 18-inch Gateway monitor with 1280 x 1024 resolution.

For the live environment, the unmanned vehicle was a Traxxas E-Maxx radio controlled car equipped with three Axis 2006 network cameras and a Netgear WGR614 wireless router (Figure 5). The vehicle was modified to mount three cameras and to improve handling and stability (see McDermott & Fisher, 2009 for a full description of the robot and environment).

Vehicle speed and orientation were controlled using a pistol-grip controller. A laptop with the following specifications was used to display video from the unmanned vehicle: a Dell Inspiron XPS M1330 and the display was a 13.3-inch WXGA screen with LED backlight.

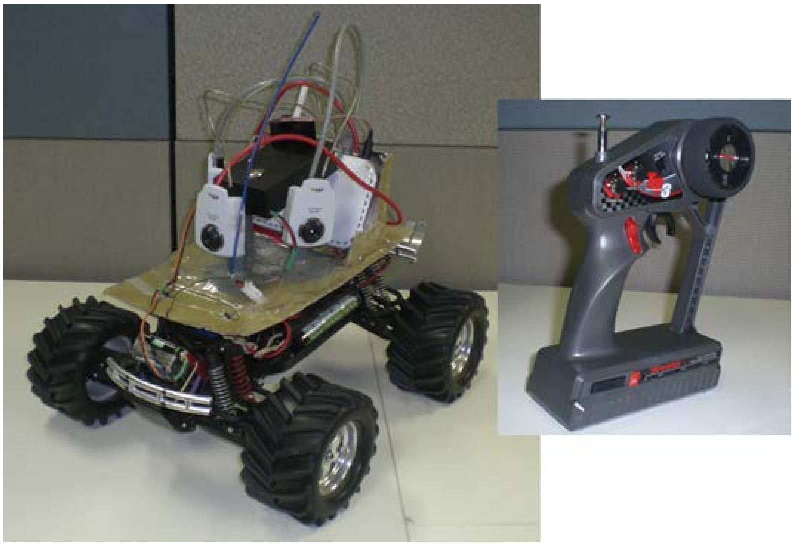

***Figure 5.* The remotely operated vehicle and controller used in the live environment.**

Stimuli

Cities. There were three cities and eight unique routes through the cities. The course consisted of two-dimensional drawings of buildings on canvas as well as three-dimensional boxes. The tanks and vehicles were 1:35 scale. The vehicles consisted of:

- A World War II-era tracked friendly tank,
- A World War II-era tracked enemy tank,
- A neutral fire engine,
- A neutral pickup truck,
- An enemy Technical (a pickup truck with a machine gun mounted in the bed), and
- A friendly enemy troop carrier.

Questionnaires. A demographic questionnaire collected basic information from each participant such as age, gender, education, and videogame experience. The Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002) was used to solicit self-ratings of spatial ability. The scale consists of 15 statements (e.g., “I very easily get lost in a new city”) to which participants rate their agreement on a scale from 1 (*strongly agree*) to 7 (*strongly disagree*). The Santa Barbara Sense of Direction Scale has been used in studies of ground robotics operators in military reconnaissance missions (Chen, 2010), the ability to learn routes with interference (Wen, Ishikawa, & Sato, 2011), and driving habits (Turano et al., 2009).

Situation Awareness (SA) sheets were used to collect participant recollections after each scenario in the live environment. The SA sheets contained a route-less map and participants were instructed to mark (1) the route taken by the unmanned vehicle, (2) location of all detected targets, and (3) affiliation of each detected target (e.g., friendly, enemy, neutral).

A post-experiment questionnaire was given verbally by the experimenter to solicit reactions to the training method, including aspects that were easy and aspects that were challenging.

Training materials. Participants viewed a PowerPoint presentation that showed the different experimental vehicles from three different angles (side, front, and back) and three different camera types (daylight simulation camera, infrared simulation camera, and live daylight camera). This resulted in one slide per asset with nine images. As an example, the side view of the neutral pickup truck is shown in Figure 6 from three different camera types. After displaying one slide per vehicle, comparison slides showed all seven vehicles from a common angle from the daylight simulation camera, allowing the user to perceive similarities and differences among the vehicles. One slide showed the seven vehicles from the side, another slide showed the vehicles from the back, and a third showed the vehicles from the front. The presentation pace was controlled so that each participant viewed each slide for the same amount of time.

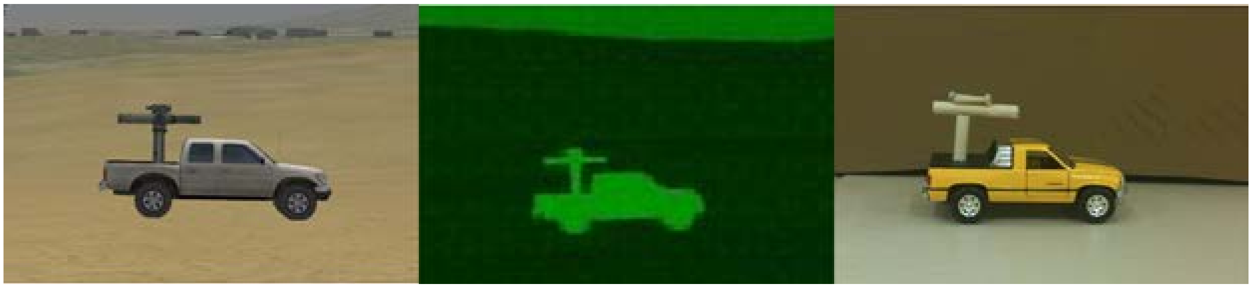

Figure 6. **Simulation, infrared, and live views of the pickup truck from the side.**

Procedure for Session 1: Simulation Training

Part-task simulation training (p), used in the p-W and the p-PW conditions. Participants completed four practice teleoperation scenarios in which they drove the unmanned vehicle along a predetermined route. In the practice scenarios, wrong turns were immediately corrected so that participants stayed on the correct route. Participants then completed a full-length teleoperation scenario until they did it to criteria (i.e., no wrong turns, within a 3.5 minute time limit). Corrective feedback regarding wrong turns and mission time was given at the conclusion of each scenario.

This general process was repeated for detection scenarios: four practice sets followed by a full-length scenario that was repeated until a criterion of 100% accuracy was reached. For detection, the participant was not teleoperating the vehicle, but observing video of the autonomous unmanned vehicle traversing the course. After each scenario, instructors used the map to point out the location of any missed targets.

Prior to identification scenarios, participants viewed identification training slides. They then completed four practice identification sets of increasing difficulty in which a vehicle was displayed and participants hit a button to classify it as friendly, enemy, or neutral. Each identification was followed by immediate corrective feedback. Following the four practice sets, participants completed a full-length identification set until they reached a criterion of 100% accuracy.

Part-task and whole-task simulation training (pw), used in the pw-W condition. Participants completed the same practice scenarios as the part-task condition: four teleoperation scenarios, four detection scenarios, and four identification sets. Then, participants completed full-length experimental scenarios. The full-length scenario integrated the three skills so that the participant simultaneously teleoperated the vehicle, detected vehicles, and identified the vehicles as friendly, enemy, or neutral. Corrective feedback was provided at the end. This included:

- Percent of vehicles correctly detected. Instructors used the map to point out the location of any missed vehicles and provided details on why the vehicles were missed (i.e., the turn was too wide and the vehicle was not in the field of view, the vehicle was only briefly in the field of view, etc.);
- Any false alarms (participants indicated there was a vehicle when there was no vehicle present) and double detections (the same vehicle detected more than once);
- Percent of vehicles correctly identified. For those incorrectly identified, instructors told participants the correct answer and context for the vehicle (i.e., in the beginning of the scenario, when you crested the hill, you identified the vehicle as Enemy. It was a Friendly Abrams Tank.); and
- Mission time.

These full-length scenarios were repeated until the participant reached criteria of no wrong turns, 100% detection accuracy, no false alarms or double detections, 100% identification accuracy, and within a 5-minute time limit.

Whole-task simulation training (w), used in the w-W condition. Participants began by viewing the identification training slides. Then participants completed four practice scenarios of increasing complexity. Each scenario involved teleoperation along predetermined routes, detection of vehicles, and identification of vehicles. These whole-task training scenarios involved the same route as the teleoperation part-task scenarios, the same vehicle locations as the part-task detection scenarios, and the same vehicles as the part-task identification scenarios. Thus, every participant drove over the same ground and saw the same vehicles. The only difference was whether they did these tasks simultaneously or separately. In the practice scenarios, wrong turns were corrected immediately so participants stayed on a correct path that included vehicles. Corrective feedback was provided at the end of the scenarios and was identical to the feedback provided in the full-length scenarios in the *part-task and whole-task simulation condition*.

After the practice scenarios, participants completed full-length experimental scenario until they completed the scenario with no missed detections, no identification errors, no wrong turns, and within a 5-minute time limit. No feedback was given during the scenario, but corrective feedback (including a demonstration on the map of any wrong turns) was provided at the end of the scenario. Figure 7 shows the paper map with the route illustrated with a dotted line. The numbers illustrate the location and type of target vehicles in the scenario. This is identical to the map the participants used except that the participant map did not include the numbered vehicle locations.

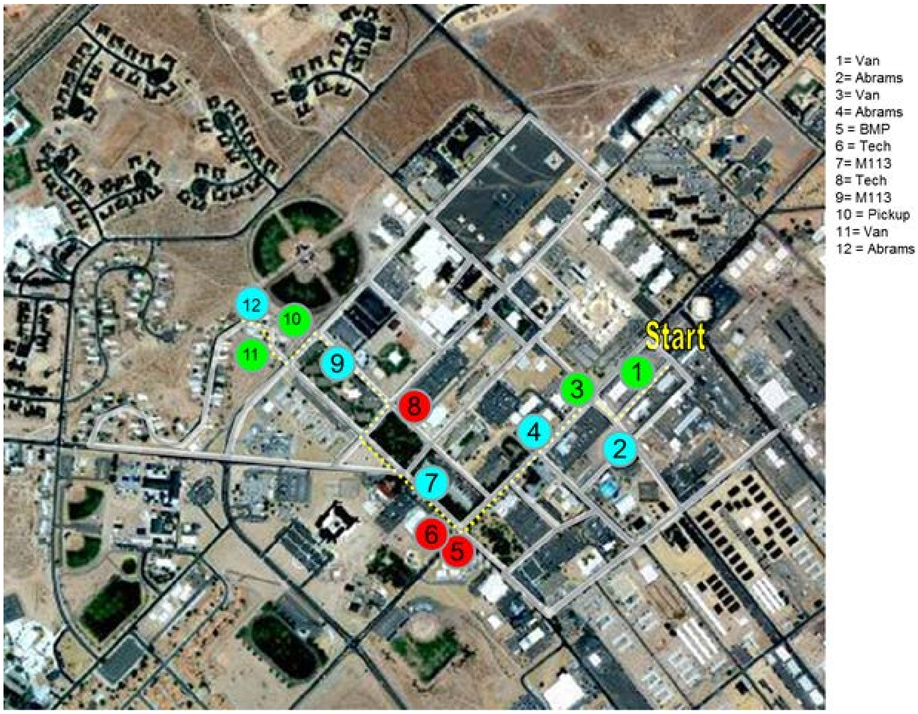

Figure 7. **SMI whole-task route with vehicle locations.**

Procedure for Session 2: Live Environment

The second session occurred 1 week after Session 1. It lasted 2 hours. Only the p-PW condition included training; the other conditions began with whole-task transfer scenarios. The p-PW condition was included in order to provide evidence and guidelines about *when* it is most useful to apply a part-task training approach. Comparing the p-PW condition to the p-W condition provided insight on whether the additional part-task training, on the same day as transfer, was worth the extra training time – both in terms of total trainee involvement time and total time on the robotic system. In other words, did providing additional part-task training on the day of transfer result in quicker achievement of criteria or did it result in less time spend on the expensive robotic system?

Part-task live training (P), used in the p-PW condition. Participants started with three practice teleoperation scenarios using an actual unmanned ground vehicle. After teleoperation, participants completed four detection scenarios and three identification sets. The detection and identification training was conducted on the same part-task training application that was used in simulation training. The only difference was that the video and static images embedded in the application were generated using the actual unmanned vehicle, instead of the simulated vehicle. After completing part-task training, participants completed whole-task transfer scenarios identical to all other conditions.

Whole-task transfer. Participants first completed a full-length integrated scenario in which they teleoperated the robot along a predetermined route, detected vehicles, and identified vehicles. Figure 8 shows a sample route with vehicle locations marked. They repeated scenarios of similar difficulty until they met the following criteria: 100% detection accuracy, no false alarms or double detections, 100% identification accuracy, no wrong turns, no collisions, and within a 5.5-minute time limit. Immediately following each scenario (before feedback was given) the map was removed and participants were given SA sheets (i.e., routeless maps) and instructed to draw their route and the location and type of vehicles encountered. SA performance did not impact whether they passed the scenario. Secondly, once they reached criteria, they completed the freeform scenario in a new city layout. If participants did not pass after eight scenarios (or within 95 minutes), they progressed automatically to the second freeform scenario. Therefore, not all participants achieved criteria. In the freeform scenario, there was no given route; the participants could drive in any route they liked. Their goal was to detect and identify all enemies in the city. They were instructed not to use the interface to note the location of friendly or neutral vehicles, as they would clutter the system. The freeform scenario concluded when the participant was satisfied that he or she had thoroughly searched the area.

Dependent Measures

The transfer performance measures were: detection accuracy (percent of targets detected), false alarms (detecting a target that was not present), double detections (detecting a target twice), wrong turns, collisions, and mission time. Because performance criteria included 100% accuracy and no collisions or wrong turns, the primary measure was the number of scenarios it took to achieve transfer criteria. In the first transfer scenario, some participants did not reach criteria by the time they needed to perform the second transfer scenario – i.e., the freeform scenario. In those cases, we adjusted the number of scenarios to equal the maximum number of scenarios (eight scenarios). This was called the adjusted scenarios to criteria. In the freeform scenario, the transfer performance measures were: time, collisions, and enemy detected.

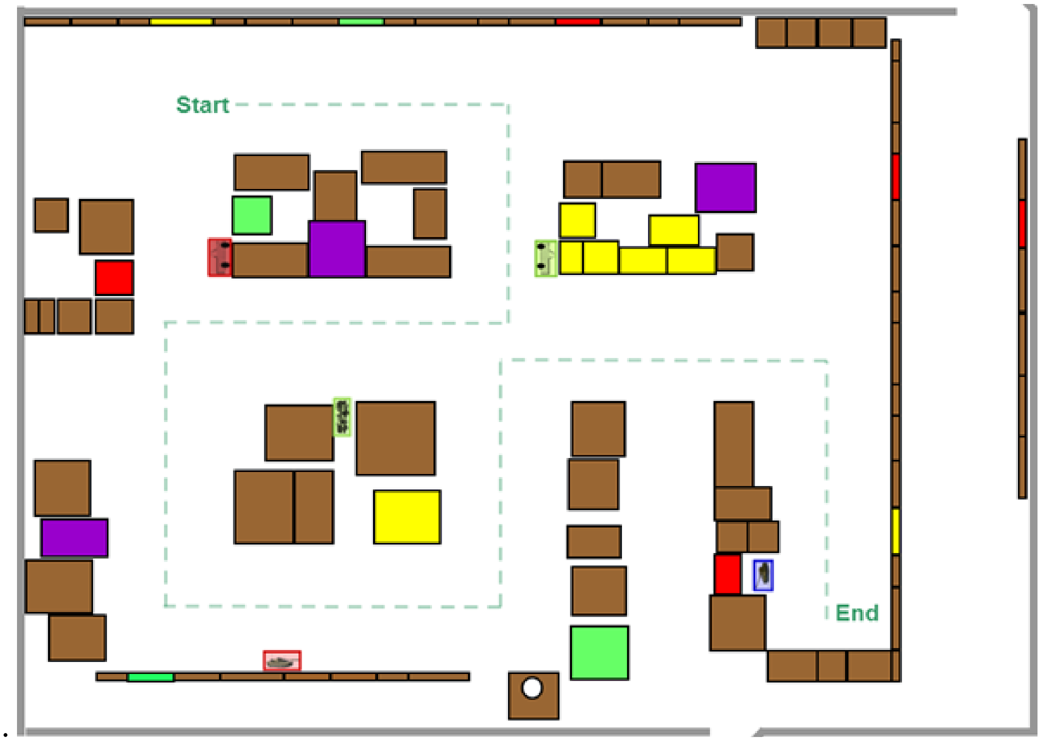

Figure 8. **Live route with vehicle locations marked.**

Situation awareness. One objective of part-task training is to reduce the cognitive effort required to manage the complexity of the training task (i.e., task components and component interactions), thereby allowing for more application of effort to learning (e.g., Salden, Paas, & Van Merriënboer, 2006). We developed an SA questionnaire to assess participants' recall of their perceptions of routes and targets detected during the scenario. Our expectation was that, consistent with cognitive load theory, participants who were more comfortable in the task would have more mental resources to devote to noting and remembering the location of vehicles and the route taken. In coding the sheets, we compared their SA responses to their perception during the scenario. For example, if they did not detect a vehicle during the scenario, a missing vehicle on the SA sheet was not considered incorrect because the vehicle was not part of their awareness. Instead, the coding scheme focused on how well the participants were able to reconstruct their experience. The SA sheets were coded for the following errors:

- Route: A deviation from the actual route taken. Each segment (i.e., straight line between two turns) was counted as one error.
- Target type: The target type differed from the target type indicated during the scenario.
- Small target location: The target was marked two buildings away or directly across the street from its true location.

- Large target location: The target was marked more than three buildings away from its true location.
- Missing: The target was not marked on the SA sheet but was detected during the scenario.
- Extra: An extra target was marked on the map that did not exist.

The errors were weighted and combined to develop an overall SA score per scenario. All errors were worth 1 point with the exception of small target errors. Because they were rather close to the actual location, they were worth only half a point. The SA sheets were coded independently by two coders. The coders met to compare scores and come to consensus for every scenario.

Results

Demographics

Across all participants, the mean hours spent playing computer/video games over the last 2 years was 8.68 hours per week (*SD* = 11.72). The mean hours per week spent playing action computer/video games in the past 6 months was 2.42 hours (*SD* = 5.57). The mean sense of direction score was 77.41 (*SD* = 12.29) on a scale from 1 to 105, with 105 being the best sense of direction. Pearson correlations were run and there were no significant correlations between age, gaming experience, and sense of direction. Independent samples *t*-tests were conducted comparing gaming experience and direction sense scores by gender. Mean differences were not significant. Finally, there were no significant correlations between any of the demographic variables and training condition.

Results from Session 1: Simulation Training

Scenarios to criteria. There were no significant correlations or mean differences between any of the four demographic variables (i.e., gender, age, computer/game experience, sense of direction) and any of the Session 1 measures. One-way analysis of variance was conducted on the number of scenarios to achieve performance criteria on mobility (collisions and wrong turns), target detection (detection accuracy, false alarms, double detections) and target identification (identification accuracy) for the part tasks (in the p-PW and p-W conditions) and on the whole task (in the pw-W and w-W conditions).

The analysis of variance on the part-task training conditions revealed a significant main effect for *mobility* part-task scenarios to criterion, $F(1, 18) = 8.04$, $p = 0.01$. Participants in the p-W condition achieved criteria in fewer scenarios than those in the p-PW condition (2.20 versus 3.20 mobility scenarios). Because there should be no difference in training for these two conditions (there were no experimental manipulations in the first session between the conditions), the finding may be due to the difference in computer gaming experience between the two groups. Although the difference in computer gaming experience was not statistically significant, at 13.30 (*SD* = 18.98) hours per week in the last two years for the p-W group and 5.05 (*SD* = 6.08) for the p-PW group, the difference was notable. An analysis of covariance showed that with the demographic covariates, individually or combined, the difference in mean scenarios to criterion for the part-task mobility was still significant at the $p < .05$ level.

The difference in mean scenarios to criterion between the w-W condition and the pw-W condition was not significant. Those in the pw-W group had a mean of 3.40 scenarios (*SD* = 2.32) to reach criterion while those in the w-W group took a mean of 4.11 scenarios (*SD* = 2.85) to reach criterion. The task combined both a segmented (detect then identify) and a fractionated (detect/identify while maintaining mobility) component. Although the two tasks are not integrated in the whole task (they do not interact, but they do interfere due to attention allocation), performing the whole task requires learning some resource timesharing skills. The non-significant difference suggests that, unlike other research which used part-task training for timesharing tasks, in this case the part-task training did not have a significant negative effect on the time required to learn the timesharing skills in the whole task.

Results from Session 2: Live Environment

First transfer scenario. One-way ANOVA for training condition yielded no significant differences in performance on the first transfer scenario for detection accuracy, identification, collisions, wrong turns, or for mission time. Independent samples *t*-tests indicated significant mean differences between males and females on number of collisions ($t(37) = 3.16$, $p < 0.01$), number of wrong turns ($t(37) = 2.85$, $p < 0.01$), and mission time ($t(37) = 2.67$, $p < 0.05$). Females were associated with poorer mobility performance (more collisions and wrong turns) and, perhaps consequently, longer mission completion times than males (see Table 7).

Table 6
Mean Performance in the First Transfer Scenario by Gender

	Collisions	Wrong Turns	Mission Time (seconds)
Female	8.80	3.87	239.60
Male	2.50	0.96	170.79
Total	4.92	2.08	197.26

Subsequent analysis of covariance with gender yielded a close to significant ($p = 0.056$) effect of training condition on number of collisions, with pairwise comparisons indicating a significant ($p = 0.05$) difference between conditions p-PW and p-W (see Table 2). This is an interesting result. Different control devices were used in the training and transfer environments. Part-task training in the live environment provided the opportunity to adapt to the new control device. In addition, if we consider number of collisions as an indicator of level of mastery of the timesharing between detection and mobility tasks then, for participants who received only part-task training in the simulation environment, it appears additional part-task training in the live environment before transfer to the whole task provided a benefit to learning the timesharing skills during the whole task. However, the additional part-task training did not benefit detection or identification skills directly, as might be expected if the benefit was due solely to the extra practice time. On the first transfer scenario of Session 2, there were no significant differences due to training condition for mission time or number of wrong turns. These results are explored further on pages 19 through 24 through examination of the learning transfer curves for each of the training conditions and skills.

Table 7
Mean Number of Collisions in the First Transfer Scenario

Condition	N	Mean	Standard Deviation
p-PW	10	2.00*	2.36
p-W	10	9.10*	10.40
pw-W	10	4.00	4.47
w-W	9	4.56	5.62
Total	39	4.92	6.74

*Indicates statistically significant difference.

Achieving criteria. While the first transfer scenario was a traditional transfer of learned skill paradigm, another important learning transfer measure is the ability to quickly get to a proficient level of performance on a new system. Twenty-nine of the 39 participants achieved transfer criteria within the allocated 95 minute time period. A chi-square test indicated there were no significant differences between training conditions in the number of participants that passed. Of the ten participants that failed to meet criteria, nine were female. This means that nine out of fifteen females failed to meet criteria and one out of twenty-four males failed to meet criteria. There was a significant negative correlation between age and meeting criteria ($r = –0.338$, $p = 0.044$). A chi-Square test between the two categorical variables (gender and ‘met criterion’) indicated a significant difference (X^2 (1) = 15.1, $p < .001$) between males (23 of 24) and females (6 of 15) in meeting performance criteria. Younger participants and male participants were more likely to meet performance criteria than older participants and female participants. Subsequent analysis of covariance did not yield significant differences due to training conditions.

Scenarios to criteria. The four training conditions were equally effective at getting 70-80% of the learners to criterion level performance within the allocated time period. For those participants who achieved performance criteria, were there differences in the number of scenarios required to achieve performance criteria based on training condition? There was no significant correlation or mean group difference between any of the four demographic variables and scenarios to achieve transfer criteria. The descriptive statistics for all conditions are shown in Table 3. Note that the p-PW condition had the fewest scenarios to criterion and the maximum in that condition is only 3 scenarios. The one-way analysis of variance yielded a significant effect for scenarios to criterion, $F(3, 25) = 4.89$, $p = 0.008$. Multiple post hoc comparisons using the Bonferroni method indicated a significant difference ($p = 0.006$) between the p-PW and the pw-

W training conditions. Participants achieved criteria significantly faster in the p-PW condition than in the pw-W condition (2.43 versus 5.92 mean scenarios to criteria). However, while this is a traditional part-task transfer paradigm (Wightman & Lintern, 1985), the additional part-task training time in the live environment (p-PW) versus whole-task training in the simulation (pw-W) must be considered in interpreting these results.

Table 8
Scenarios to Transfer Criteria for Participants Who Achieved Criteria

Condition	N	Mean	Standard Deviation	Standard Error	Minimum	Maximum
p-PW	7	2.43	.54	.20	2	3
p-W	8	4.13	1.46	.52	2	7
pw-W	7	5.29	2.14	.81	2	8
w-W	7	3.43	1.13	.43	1	4
Total	29	3.83	1.71	.32	1	8

Transfer task 2: Freeform scenario. Although transfer from teleoperation of a simulated vehicle in a simulated environment to teleoperation of a vehicle in a live environment did require adaptation of learned skills, the freeform condition provided a second adaptive transfer task because the task was different (i.e., no route provided, only note enemy vehicles). One-way analyses of variance showed no significant differences for the number of collisions, the number of enemy detected, or scenario time for the freeform scenario due to training condition. The primary criterion variables for the freeform scenario—collisions, mission time, and number of enemy detected—are shown in Table 4. Independent samples *t*-tests indicated that, as in the first transfer scenario, there were significant mean differences between males and females on number of freeform scenario collisions ($t(37) = 3.01$, $p < 0.01$) and on freeform mission time ($t(37) = 3.18$, $p < 0.01$), indicating that male participants tended to complete the freeform scenario faster and have fewer collisions than female participants. In addition, there was a significant negative correlation between the sense of direction score and the number of freeform scenario collisions; indicating that participants with a higher self-reported sense of direction tended to have fewer collisions than participants with a poorer self-reported sense of direction. Subsequent analysis of covariance did not yield significant differences in freeform mission time or number of collisions due to training condition. As would be expected, the overall mean performance of participants who had achieved transfer criteria was significantly better on the freeform scenario than participants who did not achieve criteria for number of collisions($t(37) = 2.19$, $p < 0.05$), mission time ($t(37) = 2.19$, $p < 0.05$), and identification accuracy($t(37) = -2.28$, $p < 0.05$).

Table 9
Collisions, Mission Time, and Number of Enemy Detected in the Freeform Scenario

		N	Mean	Standard Deviation	Standard Error	Minimum	Maximum
Collisions	p-PW	10	1.90	2.81	.89	0	8
	p-W	10	2.10	2.47	.78	0	8
	pw-W	10	2.40	2.12	.67	0	5
	w-W	9	1.44	2.13	.71	0	6
	Total	39	1.97	2.33	.37	0	8
Mission Time	p-PW	10	255.90	122.56	38.76	102	464
	p-W	10	247.00	102.13	32.30	151	500
	pw-W	10	285.30	81.64	25.82	146	418
	w-W	9	282.33	50.75	16.92	184	356
	Total	39	267.26	91.82	14.70	102	500
Enemy Detected	p-PW	10	2.90	.74	.23	2	4
	p-W	10	2.80	.63	.20	2	4
	pw-W	10	3.30	.68	.21	2	4
	w-W	9	3.11	.93	.31	1	4
	Total	39	3.03	.74	.12	1	4

Table 10
Mean Collisions in the Freeform Scenario for Participants Who Achieved Transfer Criteria

Condition	N	Mean	Standard Deviation
p-PW	7	.43	.54
p-W	8	2.50	2.62
pw-W	7	2.00	1.83
w-W	7	1.00	1.41
Total	29	1.52	1.90

Learning Curves for Skills

A primary concern with pure part-task training where the part tasks are performed concurrently in the whole task is that the timesharing skills are not learned during the part-task training. Analysis of the first transfer scenario results suggested some benefit to learning the timesharing skills when additional part-task training is provided in the live environment prior to the transfer task. Learning curves were plotted for individual skills to better visualize the patterns of learning over time and make qualitative comparisons between conditions. To reduce the “noise” in the learning curves, and focus separately on participants who achieved transfer criteria, we separated the participants who passed criteria prior to the freeform scenario from those who did not. In the graphs, as the scenario number increases, the number of participants attempting each scenario decreases. Therefore, to present a clearer picture of group learning over time, we assigned a perfect score for participants in scenarios after they had achieved performance criteria.

Mobility learning curves. Figure 9 shows the mean number of collisions (i.e., the mobility task) per condition for those participants who achieved transfer criteria. The p-PW participants start with the fewest collisions. They achieve zero collisions on the second scenario and do not make any collisions after that point.

On the other end of the performance spectrum are the p-W participants. They start with the most collisions and their collisions decrease in a linear fashion until the fifth scenario, in which there are no collisions. Only one individual in the p-W group who passed W criteria made any errors after Scenario 5. For the p-W group, the trend for participants who did not achieve transfer criteria is similar: initial performance is worse than other conditions and continues to be worse for the duration of the scenarios (see Figure 10).

In between these two conditions are the pw-W and the w-W conditions, which have identical curves for the first three scenarios. At Scenario 4, the w-W group asymptotes at zero collisions while the pw-W group does not reach a mean of zero collisions until the seventh scenario. Note that the maximum collisions of those who met criteria and those who failed to meet criteria are quite different (Figure 9 and Figure 10), with a maximum of 6 and 22, respectively.

The mobility curves suggest that for the two conditions of participants who only received part-task training in simulation, the p-PW group was able to learn the mobility task during the live part-task training and then master the timesharing skills after one whole-task scenario in the live environment. For the two conditions that received whole-task training in the simulation environment (pw-W and w-W), it appears that they were able to transfer the timesharing skills to benefit initial performance during transfer, but still took some time to master the individual tasks, relative to the p-W group.

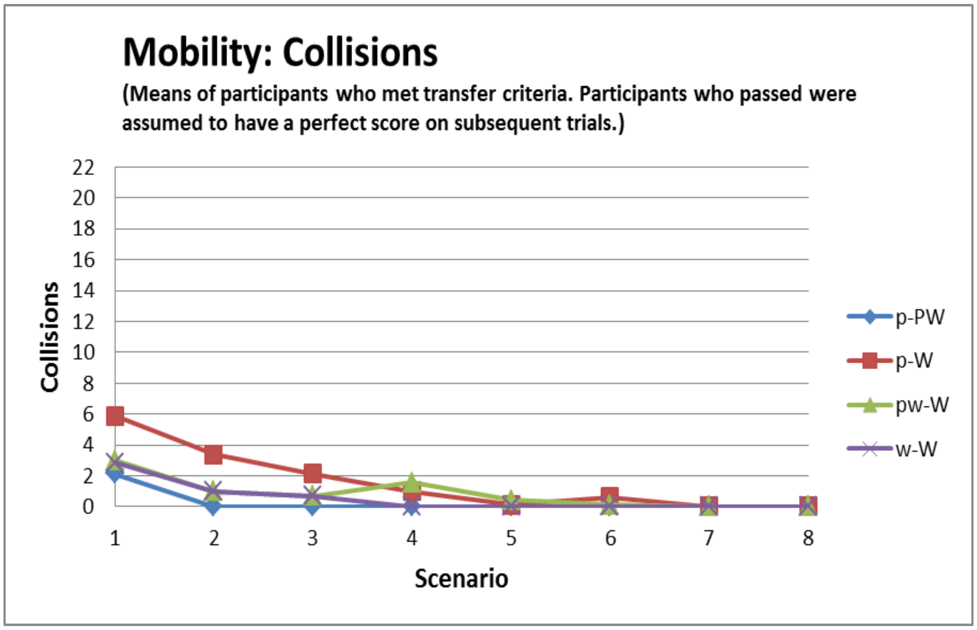

Figure 9. **Collision learning curves for those who met the transfer criteria.**

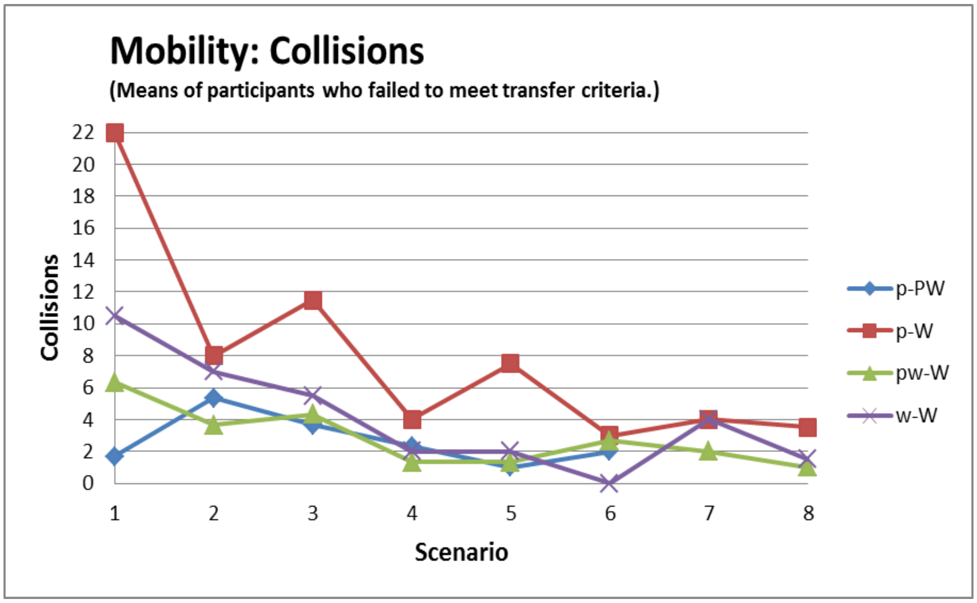

Figure 10. **Collision learning curves for those who failed to achieve transfer criteria.**

The other aspect of the mobility task is staying on the correct route. Figure 11 shows the mean number of wrong turns per condition for those participants who achieved transfer criteria. The p-PW participants start with a mean of less than one wrong turn and asymptote at zero by the second scenario. Those in the pw-W group achieve a mean of zero wrong turns by the second scenario, with the exception of one participant who makes wrong turns in the sixth scenario. Those in the w-W condition have zero wrong turns from the first scenario and only one participant makes any wrong turns (in scenario 2). Those in the p-W condition take the longest to achieve zero errors, which they do at the fourth scenario.

As seen Figure 12, the trends are not as clear for those who did not achieve transfer criteria. Those in the w-W condition have a mean of zero wrong turns by the second scenario. There was one anomaly in Scenario 7 in which a participant mixed up the start and end point halfway through the scenario and returned to the start point instead of the end point. For all conditions, errors decrease in a linear manner until Scenario 3. After that point there continue to be some increases in errors for all conditions.

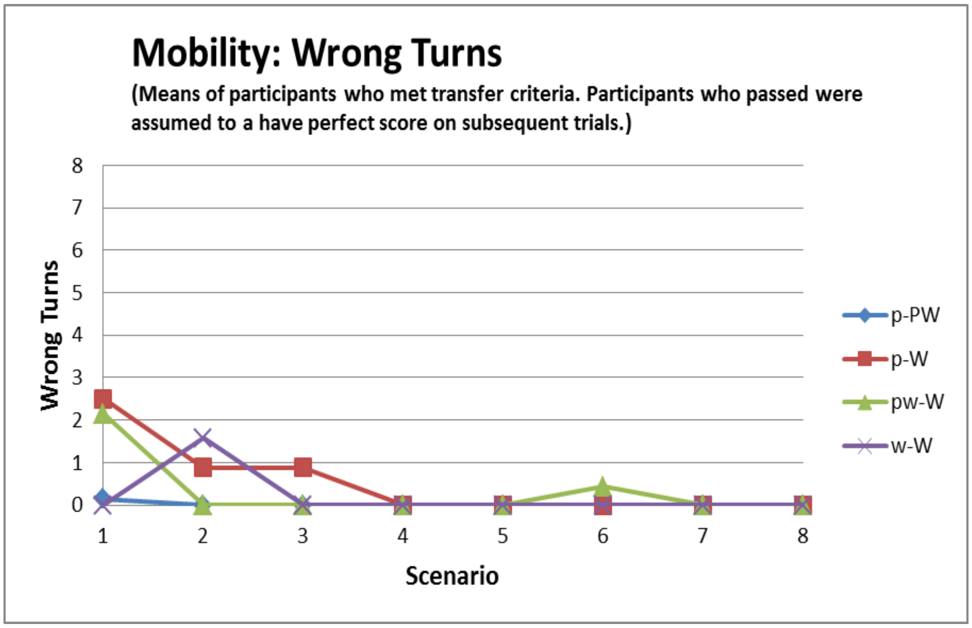

***Figure 11*. Navigation learning curves for those who met the transfer criteria.**

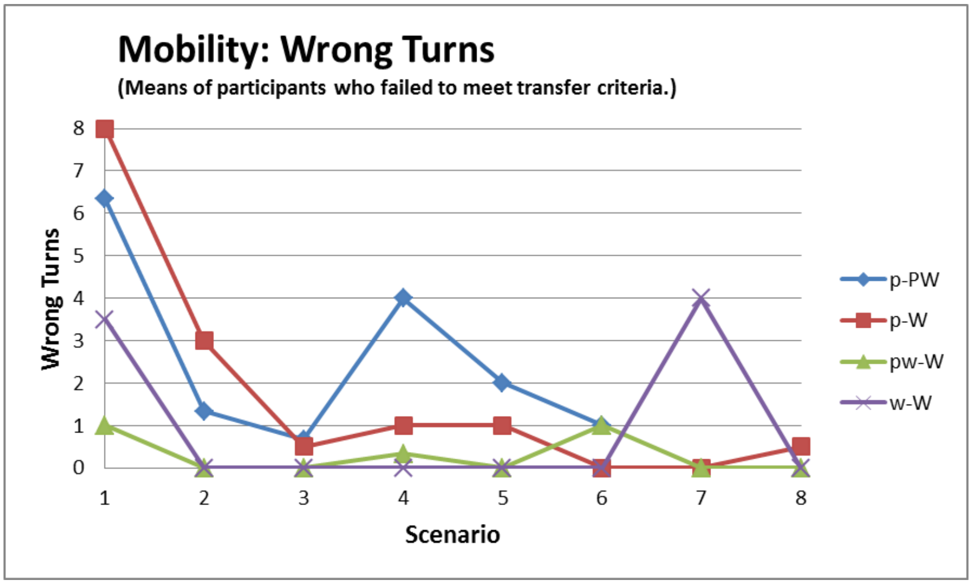

Figure 12. **Navigation learning curves for those who failed to achieve transfer criteria.**

Detection accuracy learning curves. Figure 13 shows *detection* curves for those who achieved transfer criteria. Overall, the mean performance on the first scenario is relatively high: between 77% and 89%. There do not appear to be meaningful differences between training conditions. For those who did not achieve transfer criteria, the patterns are the same shape but have a wider range and are more erratic than those who achieved criteria (see Figure 14).

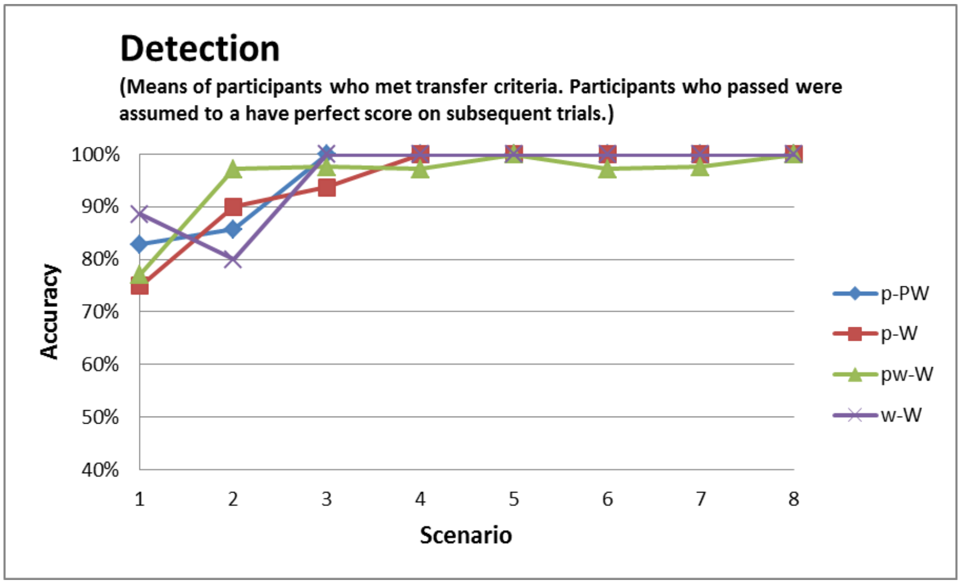

Figure 13. **Detection accuracy learning curves for those who met the transfer criteria.**

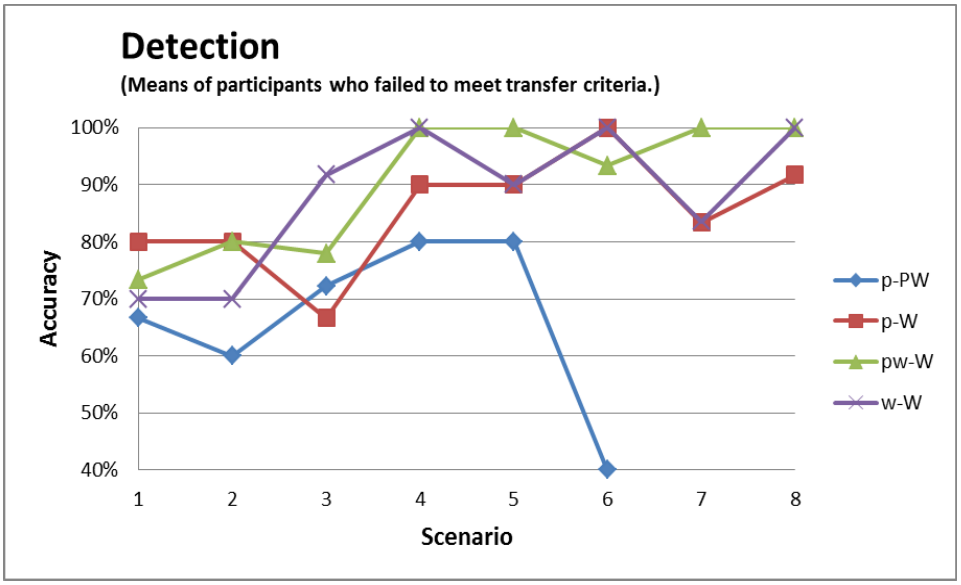

Figure 14. **Detection accuracy learning curves for those who failed to achieve transfer criteria.**

Identification accuracy learning curves. Figure 15 shows *identification* curves for those who achieved transfer criteria. Overall, the mean performance on the first scenario is relatively high: between 78% and 96%. The p-PW group stands out as having the best initial performance and being the quickest to asymptote at 100% identification accuracy. The other three groups follow a similar learning curve. As seen in Figure 16, the curves for those who failed to achieve transfer criteria are erratic. As evidenced by the fact that the learning curves almost never reach 100% accuracy, the identification requirement was a key reason that these participants did not pass transfer criteria.

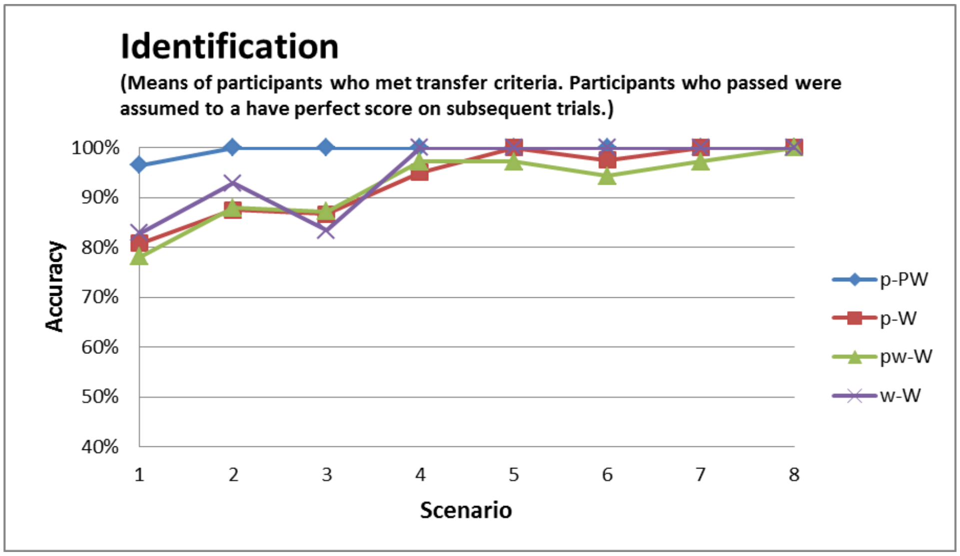

Figure 15. **Identification accuracy learning curves for those who met the transfer criteria.**

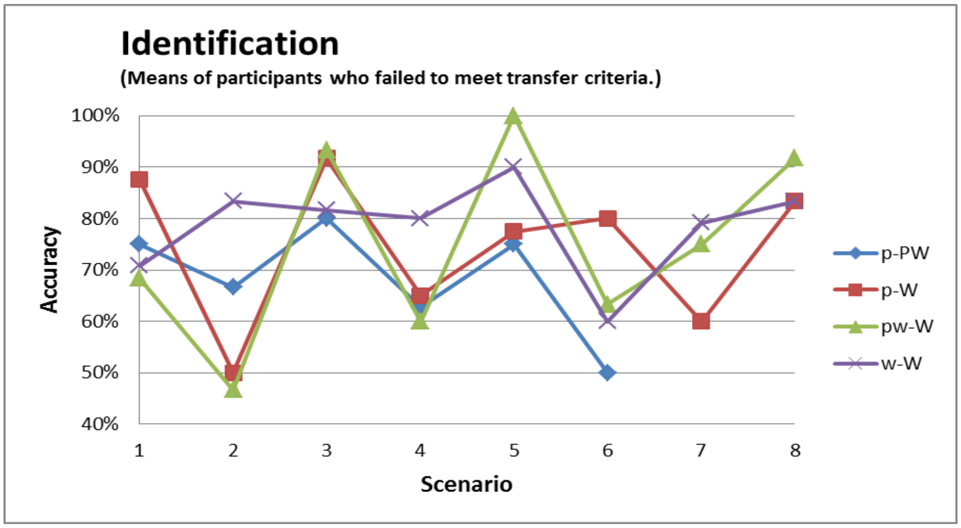

Figure 16. **Identification accuracy learning curves for those who failed to achieve transfer criteria**

Situation Awareness

The overall mean weighted SA score for the transfer scenarios by training condition is shown in Figure 17. The p-PW group had the fewest SA errors and the p-W group had the most SA errors (2.85 and 4.76 SA errors, respectively). The differences were not statistically significant.

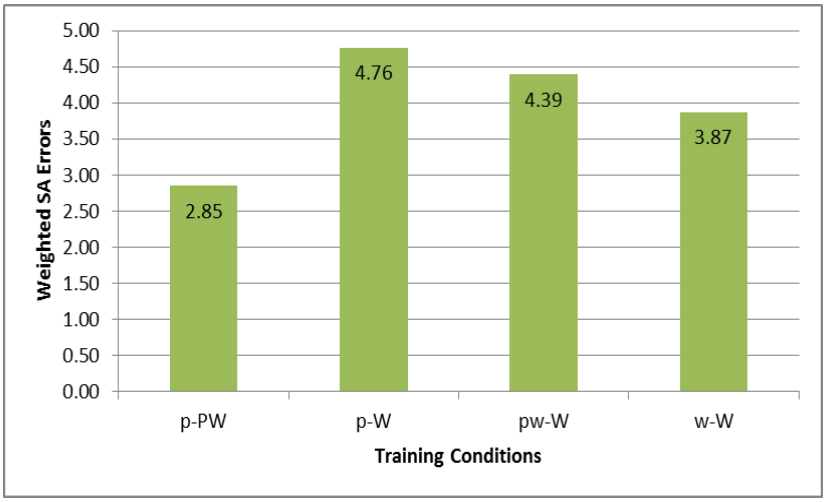

Figure 17. **Mean weighted SA score across all conditions.**

SA errors were significantly lower for those who achieved transfer criteria than for those who did not, $F(1,37) = 14.103$, $p = 0.001$. Those with better performance also had better SA, as evidenced by fewer SA errors (see Figure 18).

Figure 18. **Differences in SA score for those who passed transfer criteria and those who failed to achieve criteria.**

Figure 19 shows the SA score by training condition for the first transfer scenario and the freeform scenario for participants who achieved transfer criteria. Differences across training conditions were not significant for either first scenario or the freeform scenario. While differences between first scenario and the freeform scenario are not significant for the p-W training group, they do suggest that the p-W training group had a difficult initial transfer scenario but had more resources available for situation awareness by the freeform scenario. The additional part-task training on the robotic system in p-PW condition did not appear to provide an SA benefit during the first scenario but the trend shows that the p-PW group had the least SA errors by the freeform scenario.

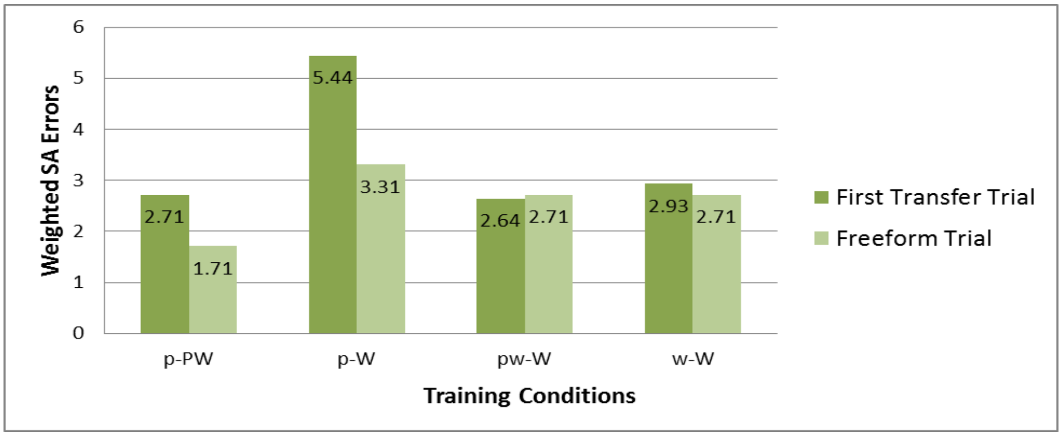

Figure 19. **SA scores for the first transfer scenario and the freeform scenario for participants who achieved transfer criteria.**

Time Spent in Training and Transfer

Participants continued working until they reached criteria; this was true for training in the simulator and transfer in the live environment. We looked at the training times to discern whether one training method resulted in more efficient learning compared to transfer times, to determine whether any savings in training time come at a cost in transfer performance. The times reported in this section represent the minutes engaged in training scenarios. It does not include the time between scenarios (i.e., time spent getting feedback or waiting for the next scenario to begin) nor the general introduction, which was consistent for all training conditions. Because there were no differences between the training conditions in the number of trainees who reached transfer performance criteria within the allocated time limit, the focus is on comparing the training time for participants who met the transfer criteria and evaluating differences between the four training conditions for total training time required to reach performance criteria in training and transfer. All data reported in this section are for participants who achieved the transfer criteria.

Session 1: Simulator training. The mean training times for the first session are shown in Figure 20 and Table 6. The time spent in part-task training is shown in dark bars and the time spent in whole-task training is shown in light bars. The p-PW and p-W training conditions have similar Session 1 training times because they experienced identical training during Session 1. Participants in the pw-W condition spent a mean of 30.29 minutes in training: 60% of their training time in part-task training and 40% in whole-task training. Participants in the w-W training condition completed training in the least amount of time (24.48 minutes) and are used as a baseline for comparison to other conditions. In comparison to the w-W condition, the pw-W condition took 24% longer (a training time of 30.29 minutes), the p-W condition took 43% longer (a training time of 35.02 minutes), and the p-PW condition took 56% longer (a total training time of 38.27 minutes).

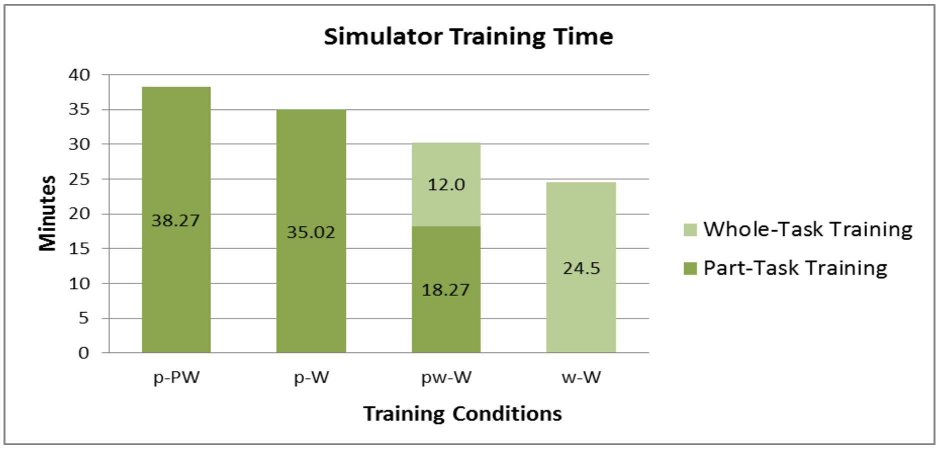

Figure 20. **Simulator training time for participants that achieved transfer criterion.**

Table 11
Mean Simulator Training Time for Participants Who Achieved Transfer Criteria

Condition	N	Mean minutes	Standard Deviation	Standard Error	Minimum	Maximum
p-PW	7	38.27	3.17	1.20	34.32	42.20
p-W	8	35.02	5.13	1.81	29.68	43.07
pw-W	7	30.29	9.82	3.71	21.77	49.45
w-W	7	24.48	9.14	3.45	14.20	41.60

The mean differences for total training time for the four training conditions was significant, $F(3, 25) = 4.80$, $p < .01$. Post hoc comparisons indicated that the difference between the w-W condition and the p-PW condition was significant ($p = .01$) and the difference between the w-W condition and the p-W condition was marginally significant ($p = .059$). As expected, whole-task training took less time than the part-task training.

Session 2: Transfer to live training. We calculated the amount of time spent in session 2. For the p-PW condition this included part-task training and transfer scenarios. For the other three conditions it only included the transfer scenarios. The p-PW part-task training consisted of a mandatory 5-minute refresher on target identification, training of detection on the part-task trainer, training of identification on the part-task trainer, and training of mobility on the robotic system. Therefore, of the 13.81 minutes spent in part-task training (see Figure 21), 5 minutes were spent reviewing target identification slides.

Participants in the w-W group spent the least amount of time in session 2, with an average of 10.91 minutes (*SD* = 5.14 minutes). The p-PW group had the highest total time in the second session, with an average of 19.96 minutes (*SD* = 2.50 minutes). A majority (69%) of their time was spent doing part-task training. The w-W condition is used as a baseline for comparison because it was the shortest time. In comparison to w-W:

- the p-PW group took roughly 1.8 times the amount of time in the transfer session;
- the pw-W group took roughly 1.5 times the amount of time in the transfer session (*M* = 16.94 minutes, *SD* = 8.51 minutes); and
- the p-W group took roughly 1.4 times the amount of time in the transfer session (*M* = 14.97 minutes, *SD* = 9.76 minutes) in training.

Analysis of variance did not indicate a significant difference in the mean training time for the four training conditions.

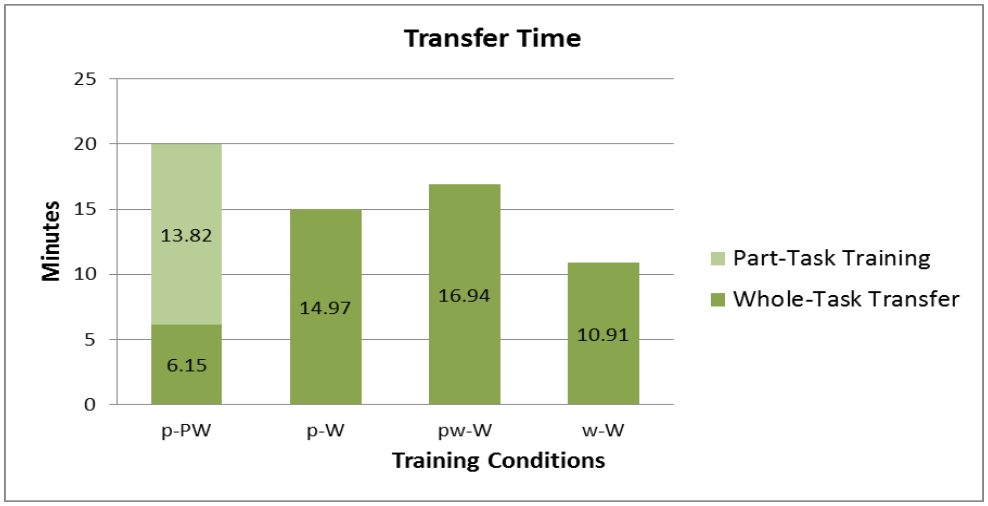

Figure 21. **Time spent in training and transfer during Session 2.**

Given the objective of reducing training costs by minimizing the amount of time spent training on expensive operational equipment, we compared the training time on the live robotic system for the four training conditions (see Figure 22). In the p-PW condition, participants used the robot to practice mobility. To practice detection and identification, participants used a simple program that presented video of imagery and recorded participant responses. In summary, the p-PW condition participants spent an average of 1.62 minutes doing part-task mobility training and 6.15 minutes doing transfer scenarios. Therefore, the p-PW participants spent a mean total of only 7.77 minutes using the robotic system. Using that number as a baseline, the p-PW condition spent 40.4% less time on the live robotic system than the w-W condition, 92.7% less time than the p-W condition, and less than the 118.0% less time than the pw-W condition. The pw-W participants spent more than twice the time on the system as the p-PW participants (9.17 mins).

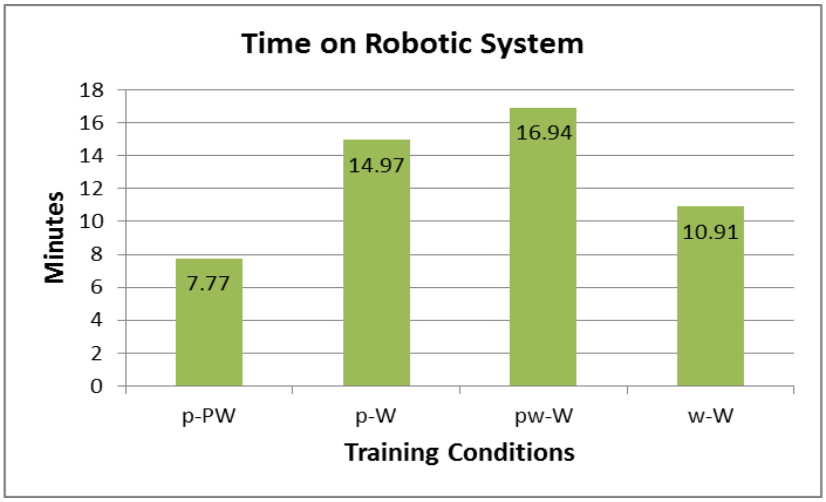

Figure 22. **Time spent using the live robotic system (part-task training or transfer scenarios).**

A common approach for measuring the transfer effectiveness of a training method is the savings in the time required to reach performance criteria in the transfer task, given the additional training (e.g., Wickens & Hollands, 2000). Percent transfer is measured as the difference between the time to reach target task criterion without prior training and the time to reach target task criterion after prior training, divided by the time to reach target task criterion without prior training. If the training results in a reduction in time required to achieve transfer criteria, there is positive transfer. Using the time to transfer criteria measures, we can estimate the transfer effectiveness of the pre-transfer part-task training in the live environment relative to the comparable group that did not receive the additional part-task training. In other words,

Transfer Effectiveness = (p-W transfer time – p-PW transfer time)/p-W transfer time

The p-pW condition is used as the basis of comparison because it only differs on one feature – whether additional part task training was conducted in the live environment on the day of the transfer tests. The p-PW condition provided a 58.9% transfer time savings relative to the p-W group, a savings of 8.82 minutes.

The ratio of the savings in transfer time (8.82 minutes) to the additional training time for the part-task training is a measure of relative transfer efficiency. Considering the total additional training time for the part-task training (13.81 minutes), the transfer efficiency ratio (time saved in transfer/additional training time) is 0.64 relative to p-W. However, if we consider transfer efficiency in terms of the cost of training on a more expensive live system, the extra training time on the live equipment (1.62 minutes) relative to the savings in transfer time on the whole system (8.82 minutes) yields a 5.45 transfer efficiency ratio for the additional part-task training relative to p-W. An accurate transfer efficiency ratio would be based on the weighted costs/benefits of each training component.

Discussion

Research Question 1 assumed that only part-task training is available prior to performance on the live system and investigated the benefit of providing additional part-task training in the transfer environment before starting the whole-task transfer. It was intended to provide insight regarding *when* part-task training was effective. Comparison of the p-W and p-PW training conditions indicates that the additional part-task training in the live environment on the same day as the transfer test did significantly reduce the number of scenarios required to achieve transfer criteria, F (1,14) = 8.42 p = .012. However, because the part-task training can be considered extra transfer scenarios, is there a transfer gain relative to the equivalent number of whole-task scenarios? That is, is the difference due only to the extra training time in the live environment or to the part-task training method? Examination of the learning curves indicates that the part-task training (provided pre-transfer in the live environment) benefited the identification and mobility tasks. The similar performance progression over scenarios on detection suggests that participants focused attention on detection at the cost of mobility. The transfer task mobility learning data, both wrong turns and collisions, suggest that learning mobility skill while timesharing with detection was difficult, even for those groups that had previous whole-task training. The additional part-task training appears to have allowed the p-PW group to focus on learning to timeshare the detection and mobility tasks during the transfer scenarios, while the p-W group had to learn detection, mobility, and timesharing skills at the same time. On average it took the p-W group seven scenarios to master the mobility task and therefore the timesharing skills, whereas the p-PW group took on average only one whole-task scenario to master the timesharing skills, as evidenced by the mobility performance. If we consider the part-task training in the live environment equivalent to three whole-task transfer scenarios, the gain in learning time is still substantial and supports the benefit of initial part-task training in the live environment to reduce the difficulty of the whole task, at least in the situation where only part-task training was available in the initial training environment. As Figures 17 and 19 show, the SA error data provide additional support suggesting that for the group receiving the part-task training in the live environment, transfer to the whole task required less focused effort, making cognitive resources available for SA, as measured by SA recall errors.

The next research questions (2 and 3) address the relative advantages of part-task versus whole-task training in the simulation environment when both options are available and transfer is directly to the whole task in the live environment. Comparison of p-W and w-W addresses Research Question 2, the relative advantages of part-task versus whole-task simulation training for transfer to the live performance environment. While the w-W group performed better than the p-W group on most first transfer scenario measures and on mean number of scenarios to achieve transfer criteria, these differences were not significant.

Comparison of pw-W and w-W addresses Research Question 3, the traditional evaluation of part-task training relative to whole-task training for training the whole task, and then for transfer to the live environment. In the second session, both of these conditions started with the whole-task transfer scenarios. Mean differences in the number of scenarios needed to reach training criteria (in the simulation training) or transfer criteria (in the live environment) were not significant. There were no significant differences in first transfer scenario measures. The difference in scenarios to reach transfer criteria was marginal ($p = .065$), in favor of w-W. The mobility learning curves and first transfer scenario mobility performance data suggest that the having only whole-task training in the simulation environment did provide some transfer benefit in the live environment a week later.

Research Question 4 addressed differences in the time required to learn the timesharing skills in the live environment for the four different combinations of part and whole-task training. In this task, the concurrent mobility and detection/identification components are separate tasks but require timesharing of attention resources for effective performance. As discussed above, based on the learning curves for detection, identification, collisions, and wrong turns, we can infer that the number of collisions is an indicator of how well the attention timesharing skill has been learned. Transfer performance in the first transfer scenario suggests that the extra part-task training in the live environment allowed this group to focus primarily on the timesharing skills during the first transfer, which helped participants learn timesharing skills faster in the whole task. Additionally, based on first scenario collisions, the groups that received whole-task training in the simulation environment appear to have transferred some timesharing skill to benefit initial transfer performance.

Research Question 5 focused on the impact of initial training and transfer training on a second transfer task that required some adapting of learned skills. The freeform scenario was designed as an additional transfer task that added an element of learner control to the task. Although the data suggest that performance for all conditions had leveled out by that point, the benefit of the additional part-task training in the live environment persisted into this second transfer task for participants who had achieved transfer criteria, with significantly fewer mean number of collisions. For those who had achieved transfer criteria, the freeform scenario likely did not require much in the way of adapting learned skills for participants with a good sense of direction.

Finally, Research Question 6 questioned whether some training methods were more (or less) beneficial for trainees with certain abilities, experience, or demographics. While individual differences did play some role in task performance, there were no interactions with training method. The number of females and males in each training condition was even with six males and 4 females in every group (with the exception of the pw-W group which only had 3 females). There were no differences by gender in training performance, only in transfer performance. There were significant correlations between gender and mobility variables and mission time on the first transfer scenario and on the freeform scenario. Male participants tended to complete the scenarios faster and have fewer collisions than female participants. Younger participants and male participants were more likely to meet transfer performance criteria. Nine out of the ten participants who did not achieve transfer criteria were female. However, with the exception of collisions on the freeform scenario, we found no correlation between sense of direction and transfer performance. This is surprising because there is a range of research reporting significant gender differences in spatial abilities, with males performing better on spatial tasks than females (Goettl & Shute, 1996). Perhaps in the present experiment, the lack of correlation could be due to the fact that performance included a substantial perceptual component (i.e., detection, identification) that was not impacted by sense of direction. An objective measure of spatial ability, as opposed to a subjective measure of perceived sense of direction, may have been correlated with performance.

Conclusions

This experiment helped to address some research questions, identified through review and meta-analyses of part-task training research which were conducted as part of the larger ongoing project, in the context of an Army-relevant task domain and training conditions. The primary focus of the research is on effective use of simulation-based training. If you have a simulation in a classroom environment, what is the best way to use it to promote transfer of learned skills? The training conditions therefore included: a simulation environment for training, transfer to a live performance environment, and delay of 1 week between training and transfer. Our research questions were not designed to address the effectiveness of training on the simulator versus doing all training in the transfer environment.

While the training method used in the simulation environment (pure part-task, whole-task, or part followed by whole-task) did not have a significantly differential impact on transfer to the whole task in the live environment, performance trends generally favored the whole-task simulator training. In addition, of the three groups receiving only whole-task training during transfer, the group receiving only whole-task training in the simulator also spent less time in reaching transfer criteria. Based on the results of this experiment, we conclude that, for tasks similar to robotics control involving teleoperation and concurrent target detection, whole-task training in the simulation environment may provide a small transfer performance and efficiency benefit over pure part-task training, but it is important to note that these observed differences were not statistically significant in the present research. Without comparison to a live whole-task control condition baseline, no conclusion can be made about the overall transfer effectiveness of part-task (or whole-task) training in the simulation environment.

However, the *additional* part-task training in the live environment provided a significant benefit to performance on the whole task, as compared to when only part-task training was used in the simulation environment. Even though the part-task training did not provide an opportunity to learn concurrent task timesharing skills, the transfer performance curves suggest that it provided the conditions for quickly acquiring the timesharing skills during whole-task transfer. This may be unique to concurrent component tasks that require attention management but not interaction management in the whole task. It is important to note a week had elapsed since the first training session and the p-PW condition was the only condition to receive any training on the day of the transfer tests. This was extra training. The other three conditions started directly with transfer tests. The p-PW trainees spend longer overall in session 2 because they were doing part-task training as well as the transfer tests. However, the part-task training in the live environment was more efficient in terms of training time on the live robotic system. In this experiment, participants spent a mean of 7.77 minutes on the more expensive robotic system. This includes both the part-task mobility scenarios and the whole-task scenarios, both of which used the robotic system. This is in comparison to those who went directly to whole-task transfer scenarios and spent a mean of 14.27 minutes on the robotic system.

This research focused on part tasks that do not interact but are performed concurrently in the whole task. This is a common task type represented in military domains but underrepresented in the part-task training research literature. The results of this experiment suggest that for concurrent but separate tasks, part-task training in the transfer environment can result in effective and efficient transfer to the whole task.

Given these conclusions additional experimentation on this task is warranted. In particular, a live whole-task control would provide a baseline for interpreting the value of the simulation-based training. Does the pure part task in the simulation provide a positive, negative, or neutral impact on training relative to no prior simulation-based training? In addition, the live whole-task control would provide a check on the transfer efficiency of the live part-task training. Systematic comparison of the part-task methodology, with increasing timesharing demands and different types of concurrent tasks, would provide additional evidence for the scope and reliability of this effect.

References

Carolan, T., Belanich, J., McDermott, P., Hutchins, S., & Wickens, C. (2011). Investigating the impact of training on performance. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL.

Chen, J. Y. C. (2010). Effects of operator spatial ability on uav-guided ground navigation. *Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction,* 139–140. ACM, New York, NY.

Frederiksen, J. R., & White, B. Y. (1989). An approach to training based on principled task decomposition. *Acta Psychologica, 71,* 89–146.

Goettl, B. P., & Shute, V. J. (1996). Analysis of part-task training using the backward-transfer technique. *Journal of Experimental Psychology: Applied*, *2*, 227–249.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica*, *71*, 147–177.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30,* 425–447.

Lintern, G. (1991). Instructional strategies. In J. Morrison (Ed.), *Training for Performance: Principles of Applied Human Learning* (pp. 167–191). New York: Wiley.

Lintern, G., & Wickens, C. D. (1991). Issues for acquisition in transfer of timesharing and dual-task skills. In D. Damos (Ed.), *Multiple-Task Performance* (pp. 123–138). London: Taylor & Francis.

Mané, A., Adams, J. A., & Donchin, E. (1989). Adaptive and part whole training in the acquisition of a complex perceptual motor skill. *Acta Psychologica*, *71*, 179–196.

McDermott, P. L., & Fisher, A. (2009). The tradeoff of frame rate and resolution in a route clearing task: Implications for human-robot interaction. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.

Paas, F., & van Gog, T. (2009). Principles for designing effective and efficient training for complex cognitive skills. In F. Durso (Ed.), *Reviews of human factors & ergonomics, Vol 5* (pp. 166–194). Santa Monica, CA: Human Factors.

Salden, R. J. C. M., Paas, F., & Van Merriënboer, J. J. G. (2006). A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior, 22,* 321–333.

Turano, K. A., Munoz, B., Hassan, S. E., Duncan, D. D., Gower, E. W., Roche, K. B., Keay, L., Munro, C. A., & West, S. K. (2009). Poor sense of direction is associated with constricted driving space in older drivers. *The Journals of Gerontology Series B, The Journal of Gerontology: Psychological Sciences and The Journal of Gerontology: Social Sciences*, 64B: 348–355.

Wen, W., Ishikawa, T., & Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. *Applied Cognitive Psychology, 25*, 654–662.

Wickens, C. D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Part-task training and increasing difficulty training strategies: A meta-analysis approach. *Human Factors, 55,* 461-470.

Wightman, D., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27,* 267–284.