

## **Application of Worked Examples to Unmanned Vehicle Route Planning**

**Patricia L. McDermott, Thomas F. Carolan, Mark R. Gronowski**  
**Alion Science and Technology**  
**Boulder, CO**  
**PMcDermott; TCarolan; MGrönowski@alionscience.com**

### **ABSTRACT**

This paper describes an experiment and results from a related meta-analysis that investigated the efficacy of worked examples for improving learning and transfer performance. Worked examples are designed to demonstrate the correct steps to take in a problem-solving process. Preventing errors during training can reduce the difficulty of the task and the associated demands on the learner. Meta-analysis results suggest that worked examples may benefit transfer effectiveness for lower complexity problems and near transfer tasks. We present results from an experiment that used worked examples to train unmanned vehicle route planning. This is an example of a complex decision making task coupled with procedural data entry on a digital system. Transfer tasks were used to evaluate the effectiveness of the worked examples in supporting 1) transfer from structured training tasks to problems requiring more inferential reasoning and 2) transfer from paper-based training to performance in a simulation environment. Results suggest that worked examples did not provide a transfer benefit, either in terms of plan content or plan sophistication. The implications of worked examples and training media are discussed as well as the use of the Tower of Hanoi puzzle as a measure of problem-solving aptitude.

### **ABOUT THE AUTHORS**

**Patricia L. McDermott** is a Lead Human Factors Engineer and Program Manager at Alion Science and Technology in Boulder, CO. At Wright State University she earned degrees in both Psychology and Human Factors Engineering. Ms. McDermott uses this blend of psychology and engineering perspectives in the study of human-robot interaction and training to support decision making and situation awareness.

**Tom Carolan** is a Senior Scientist and Program Manager with Alion Science and Technology. He has been involved in research related to human performance measurement and training effectiveness in military environments for the past 20 years. He has a Ph.D. in experimental psychology from the University of Connecticut.

**Mark Gronowski** is an Associate Cognitive Scientist at Alion Science and Technology. He graduated with a degree in Psychology from the University of Colorado in Boulder, CO.

## **Application of Worked Examples to Unmanned Vehicle Route Planning**

**Patricia L. McDermott, Thomas F. Carolan, Mark R. Gronowski**  
**Alion Science and Technology**  
**Boulder, CO**  
**PMcDermott; TCarolan; MGrnowski@alionscience.com**

### **INTRODUCTION**

Many of the tasks that a Soldier needs to learn in this digital age go far beyond simple procedural tasks. For example, consider the soldier who must learn to operate a digital system for controlling unmanned vehicles. He or she must learn the mechanisms for controlling and monitoring vehicles as well as dealing with information collected and communicating the implications of those findings. Although the tasks may sound procedural, each has a cognitive component such as recognizing when a vehicle is having trouble with terrain before the vehicle becomes stuck. Even the “straightforward” task of routing a vehicle from point A to point B is full of decisions. Which vehicle is the most appropriate given the mission and the vehicle characteristics? Are specific parameters required such as altitude minimums or speed maximums? What are the current goals regarding speed and stealth? In order to achieve those goals does it make more sense to direct the vehicle under cover, cross country, or along roads? How should terrain obstacles of forests, ridges, and rivers be handled?

In training such a complex task there are various methods to manage training difficulty to make the task more tractable and ultimately improve the transfer of skills to the field. The methods for managing training difficulty range from exploratory learning to part task training to error prevention methods, to name a few. This research focuses on the ‘worked examples’ strategy which is a type of error prevention strategy that provides the correct steps needed in a problem solving process. Typically, this strategy involves the presentation and/or demonstration of a fully or partially worked problem including all solution steps and final answer (Renkl, Stark, Gruber, & Mandl, 1998).

Preventing unnecessary errors during training reduces the difficulty of the task and the associated demands on the learner’s limited cognitive capacity. In general, the research literature supports the effectiveness of worked example strategies for novice learning and structured problem-solving, compared to traditional problem solving (Sweller, van Merriënboer, & Paas, 1998; Van

Gog & Rummel, 2010). There is some evidence that providing worked examples may be ineffective or detrimental for learners with enough prior experience to complete the task successfully on their own (e.g., Kalyuga, 2007; Van Gog & Rummel, 2010). The research is mixed on the effectiveness of the worked example method for far (adaptive) transfer tasks. Some research suggests that for novel problems requiring the generation of a new, creative problem-solving technique, the worked example effect tends to disappear (Sweller et al., 1998). Other research supports the effectiveness for far transfer tasks (Atkinson, Derry, Renkl, & Wortham, 2000). Worked examples are most often used as a training strategy for learning problem-solving skills, especially within well-structured domains (e.g., physics, computer programming, and mathematics).

In a recent meta-analysis, conducted as part of this project (Hutchins et al, 2012 submitted), the worked examples method was analyzed as a type of error prevention strategy. This meta-analysis focused on studies of training transfer. Two complementary effect size measures were used in the meta-analysis process, transfer ratio and Hedges’ *g*. We present only the Hedges’ *g* results here to simplify the discussion. The literature search and selection phase yielded 23 studies using the worked example strategy with a total of 69 data points that were included in the Hedges’ *g* effect size analysis.

The primary research question addressed the benefit of using worked examples as an error prevention strategy. Using Hedges’ *g*, and a fixed effect model, the overall effect size (0.18) indicates a significant transfer benefit for using worked examples when compared to a control condition that did not use worked examples. Of these 69 data points, 50 involved only a worked example technique, the others used worked examples in conjunction with other error prevention techniques such as prompting. Surprisingly, worked examples without prompting yields a larger effect size (0.31) than the overall analysis while worked examples with prompting yields a small non-significant effect (0.03). Follow-on analysis indicated that when the prompts

had a procedural emphasis there was a significant cost to performance relative to the control condition (-0.26) but when the emphasis for prompts was on principles the benefit to performance was maintained (0.16).

For the 'type of transfer' moderator, worked examples were effective when the transfer problem was identical (0.23) or similar (0.48) to the training problems, and when the transfer problem involved a new problem type (0.33) but not when transfer was to a more complex problem (-0.07). There were few studies where ability or experience was addressed as moderating factors. For the single research study that manipulated ability, the effect of worked examples was stronger for low ability (0.18) than for high ability (0.02) trainees. How worked examples are used in training may also impact the potential transfer benefit. The meta-analysis indicated that worked examples are more effective when presented in the instruction phase (0.83) than when used during the practice phase (0.10) or during both phases (0.15). This suggests worked examples may be more effective when used to provide an initial problem schema than when used to scaffold practice. These meta-analysis results start to provide some quantitative evidence for conditions under which worked examples may be expected to benefit transfer effectiveness – lower complexity problems, near transfer tasks – and the potential size of the benefit.

The meta-analysis results point to a number of areas for further research relevant to military training needs. One area is the application of worked examples to a broader set of problem types. In the experiment described in this paper, we extended the 'worked examples' method to training complex decision making skills in the form of unmanned vehicle route planning problems. We investigated their effectiveness as an error prevention strategy for complex planning tasks. The route planning task is a type of problem solving task that can vary in terms of the degree of structure as scenario conditions evolve from one correct solution to more than one acceptable solution.

Planning problems are often less structured as complexity increases to include interactive relations between elements and dynamic events that require replanning. In addition there are conflicting findings on the value of worked examples for more adaptive transfer and more complex tasks. In this experiment transfer tasks are used to evaluate the effectiveness of worked examples in supporting transfer from structured training tasks to less structured planning problems that require complex and inferential reasoning. The suggestion based on the meta-analysis

results is that using worked examples will not be as effective, relative to a no worked examples control, when the transfer task increases in complexity relative to the training task. Finally, we evaluate the effectiveness of using a paper-based map environment for training the route planning task compared to using a simulation environment. This is somewhat analogous to comparing training on only the conceptual aspects of the task to training both the conceptual and procedural (digital) skills involved in implementing the task in the simulation environment. In other words, is it more useful to learn planning in conjunction with the digital skills or to first learn planning via paper and then transfer to the digital system?

These research questions lead to the following experimental hypotheses.

- We expect that consistent with the results from the meta-analysis, worked examples will provide a transfer benefit when the transfer task is close to the complexity level of the training task but will not provide a benefit when the transfer task is of greater complexity than the training task.
- We expect that training, using paper maps, on conceptual aspects of the planning task will not significantly reduce transfer to the simulation environment, for individuals with prior experience on the procedural aspects of the simulation environment.
- Based on the worked example data on procedural and principal prompts, we hypothesize that worked examples will have less benefit to transfer than no worked examples for paper-based training and more benefit for simulation-based training.
- Finally, we expect that worked examples will have less benefit to higher-ability performers than to lower-ability performers.

## **METHOD**

The objective was to research the impact of worked examples and the learning media on the ability to adapt to a more complex decision-making scenario. We varied whether participants had access to a worked example and whether they learned on paper or on a digital system. The transfer was done using a more complex scenario on the digital system.

### **Experimental Task**

Participants learned to make plans in response to an Operations Order (OPORD). The OPORD described the situation and the mission, including the type of information that needed to be collected in each

geographical area. Participants decided which assets were most appropriate and how to employ them (i.e., send them on a route that uses roads, avoids hills, and flies an orbit around the area of interest). They used unmanned assets to conduct reconnaissance on Named Areas of Interest (NAIs).

This task was chosen because it is a non-trivial decision making task. There was no clear “one right answer” although there were unsuitable answers. After each scenario participants were asked to describe the information they considered in making their decisions. Thus, the focus was not just on choosing assets but on understanding the information and strategies used to make the decision.

As a proxy for subject matter expertise, the participants were given a manual of vehicle characteristics. The vehicle capabilities were broken down into the seven main considerations of METT-TC: Mission, Enemy, Terrain, Time, Troops available, and Civilians. The OPORD did not list the information using those headings but all the information was present. The manual was used to understand the advantages and disadvantages of the different assets for various situations.

Participants had information on the following notional assets:

1. Unmanned Aerial Vehicle (UAV), a small fixed wing unmanned aerial vehicle with no active defense that is hard to spot.
2. Predator, a large fixed wing unmanned aerial vehicle that has an extended range.
3. Goldeneye, a small aerial vehicle that does vertical take-off and landing and has a limited range.
4. RMax, a large unmanned aerial vehicle that does vertical take-off and landing, flies high, and is capable of lethal suppression of enemy.
5. Small Unmanned Ground Vehicle (SUGV), a maneuverable ground vehicle with limited range.
6. XUV, a mid-sized unmanned ground vehicle that is capable of entering smaller areas, is quiet, and is capable of enemy suppression.
7. Stryker, a large unmanned ground vehicle that is equal to a large manned vehicle in terms of range and capability.
8. Unmanned Ground Sensor (UGS) system, a non-mobile system that detects sounds and vibrations, is quick to set up, and hard to spot.
9. Remote Battlefield Sensor System (REMBASS), a complex non-mobile unmanned ground sensor system that captures images and sound and requires extensive setup.

A limited number of vehicles (2-5) were available in each scenario.

### Experimental Design

The experiment was a 2x2 between-subjects design (see Figure 1). Aid referred to whether the participant had access to a worked example (WE). Media referred to whether initial training was done on paper or on the digital system, the Soldier Machine Interface (SMI). Crossing these two variables resulted in four conditions:

- Worked Example Paper
- Worked Example SMI
- No Worked Example Paper
- No Worked Example SMI

Participants were randomly assigned to a condition. Each participant completed 6 scenarios in the same order.

|     |                   | Media       |           |
|-----|-------------------|-------------|-----------|
|     |                   | Paper       | SMI       |
| Aid | Worked Example    | Paper WE    | SMI WE    |
|     | No Worked Example | Paper no WE | SMI no WE |

**Figure 1. Experimental design crosses the variables of Media and Worked Examples.**

### Participants

Thirty-two volunteers completed the experiment, eight per condition. This included 25 men and 7 women, ranging in age from 18 to 38 years ( $M = 25.41$ ,  $SD = 6.45$ ). Twenty participants had a college degree or were currently working on their undergraduate degree; twelve participants did not have a college degree.

All participants had completed a previous experiment in which they spent an average of 4.5 hours learning and being evaluated on procedures to operate the SMI (Carolan, Hutchins, McDermott, & Wickens, 2011). This gave us the advantage of a civilian population who was familiar with both the SMI system as well as Army terminology such as *squad*, *platoon*, *reconnaissance*, and *phase line*. Participants had been exposed to all the current experiment tasks in the previous experiment with the exception of choosing an

asset for a plan. In the previous experiment, they were always told which asset to use. Participants were recruited from the list of seventy-two previous participants. They were compensated \$125 for their time in the 4.5 hour experiment.

### Apparatus and Stimuli

The SMI, developed by General Dynamics Robotics Systems, was used as the experimental testbed (Figure 2). The SMI employs a networked user interface for controlling unmanned air and ground assets. The SMI integrates mapping, asset control, route planning and communications functions and supports target detection, route reconnaissance, and autonomous navigation.



Figure 2. Soldier Machine Interface

Six paper OPORDs were developed by Army subject matter experts. The OPORDs contained information about the:

- Situation (current time, terrain, weather, most likely enemy course of action and location, adjacent friendly unit activities),
- Mission (unmanned assets available, task, purpose, deadlines),
- Execution (likelihood of enemy in different areas, mission timeline, specific data collection needs in different areas, any specific instructions, coordinating instructions, rules of engagement)
- Service Support (logistics update, special weapon capability), and
- Command and Signal (command succession, signal frequencies, call signs)

### Questionnaires

Demographic data regarding age, gender, education, and computer experience was reused from the previous experiment. 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.

### Asset manual

The asset manual contained details regarding the unmanned assets. This included a description of the asset, its capabilities, and information about how suitable it was for different conditions – organized by the Army’s METT-TC (Mission, Enemy, Terrain, Time, Troops, and Civilians). Asset Suitability was ranked as preferred, acceptable, and unsuitable.

### Outline of basic steps

The outline was a half-page document that listed the four major steps involved in creating a plan.

### Worked Examples

One full worked example included an OPORD in which critical information was highlighted and annotated. For example, “Logistics supply is two days behind” was highlighted with the annotation “Minimal logistics.” An accompanying document gave basic guidelines for analyzing the OPORD and walked the participant through the steps involved in creating a plan. For example, the guidelines explain that based on an understanding of the OPORD, the Goldeneye is not suitable because it is not good at following moving enemy. A map showed a solution including the assets used for each area, the route the asset took to get there, whether the route used or avoided roads and hills, and the orbit if it was an aerial asset. A partial worked example was included for the first practice scenario. This partial worked example had seven annotations, one of which pointed out that the small unmanned ground vehicle was unsuitable for both target areas because they were a long range from the unit. The first page of the worked example OPORD is shown in Figure 3.

### SMI refresher

A refresher on how to perform tasks on the SMI was available to participants in the SMI conditions. This was printed PowerPoint presentation that depicted how to accomplish the required tasks (i.e., setting waypoints, placing a graphic on the map).

Procedure

**OPORD NAME:** SI- Coyote Rampart Your platoon Your platoon See Mission time below  
**TASK ORGANIZATION** 2/B/1 CAB: 1x GoldenEye, 1x Predator, 1x Stryker Current time: 1230 25 OCT 2011

**SITUATION**

**Environment**  
**Terrain Description** The AOI is primarily rural mountainous terrain with patches of foliage and cover.

**Effects**  
**Obstacles** No anticipated effects on unmanned vehicles  
**Ave of Approach** No anticipated effects on unmanned vehicles  
**Key Terrain** Limited key terrain for unmanned ground vehicles  
**Obser/Fields Fire** Limited foliage and cover, some folds in terrain limit fields of view  
**Cover/Conceal** Limited foliage and cover, some folds in terrain limit fields of view

**Weather Description** We expect seasonal temperature to be between the low 50s to the high 70s.  
 High winds across the entire area are expected.

**Battlefield effects:** Thermal Sights      Smoke/Dust

**Enemy**  
**Enemy Most Likely COA** Location: vic 11S NJ9630

Enemy most likely using area for vehicular transit to other consolidation locations, and therefore enemy most likely COA is to use speed and stealth to move through the area while minimizing contact with any forces or hostile civilians.

**Friendly**

| Adjacent Units: Who? | Location  | Mission        |
|----------------------|-----------|----------------|
| 1st platoon          | NJ 283251 | Local security |
| 3d platoon           | NJ 291957 | Local security |

**MISSION**

**Task:** provides reconnaissance and surveillance of enemy activities w/in NAIs W&Z

**Purpose:** in order to detect enemy movements through the area

**Time:** NLT 0910 26 OCT 2011

Maneuver Graphics and Area of Operations

NAI 1 & NAI2 are both at Long Range

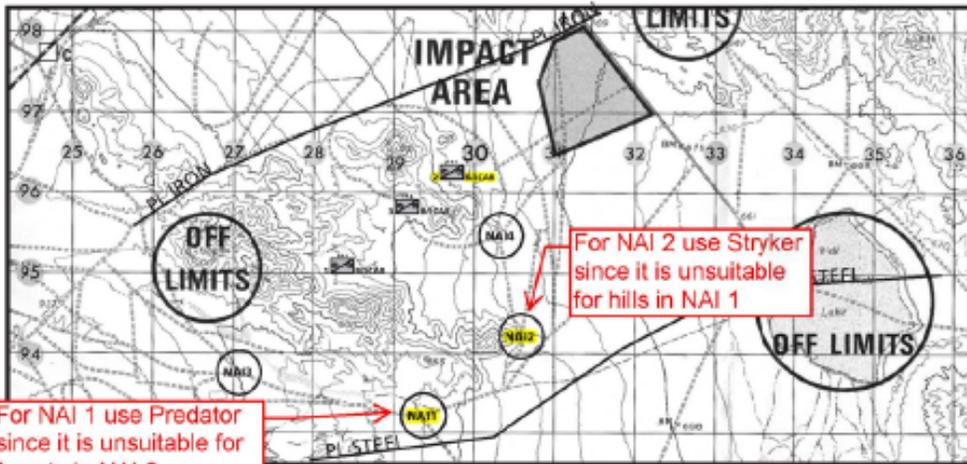


Figure 3. Excerpt of the annotated OPORD that was part of the worked example package

### First Session: Initial Training

Instructor interaction with participants was minimized in order to maintain consistency between participants. The instructor described the task and the resources available. Those in the non-worked example conditions had access to the Asset Manual and the Outline of Basic Steps. Those in the worked example conditions also had access to the full worked example. Participants read over the materials and let the instructor know when they were ready to start the first scenario. The instructor emphasized that there was not one right answer but that the focus was on considering as many factors as possible when making their decisions. Participants were told that the instructor would ask questions about the information they considered when creating the plan. Half the participants completed the plans on paper; half input the plans into the SMI.

Participants completed four scenarios as follows:

1. Practice 1: low difficulty, no time limit
2. Practice 2: low difficulty, no time limit
3. Timed 1: medium difficulty, time limit
4. Conceptual Near Transfer: timed, more challenging than Timed 1 because two assets were suitable for an area and participants had to determine how to prioritize the requirements.

Those in the worked example condition had a partial worked example for Scenario 1. They had access to the full and the partial worked example while completing scenarios 1-3. No participants had access to the worked examples in the near transfer scenario.

### Transfer Session: SMI

The second session occurred one week after the first session and lasted ninety minutes. All participants completed two scenarios on the SMI and had access to the same materials: the Basic Outline of Steps, the Asset Manual, and the SMI refresher. The scenarios were more challenging than the scenarios in the first session. Participants had to incorporate stealth, smoke, altitude, and speed into their plans as appropriate. The first exercise (FT1) evaluated the transfer from paper to the SMI. The second far transfer exercise (FT2) was evaluated a conceptual transfer to a more complex task with new problem events. The experiment concluded with a post experiment questionnaire.

### Dependent Measures

There was not one right answer or plan for each scenario. Therefore, four dependent measures were created to allow quantitative comparisons between plans.

Asset selection is a measure of plan correctness. Acceptable assets that contained no unsuitable characteristics were worth 2 points. An unsuitable asset scored zero points. It is reported as a percentage of the optimal asset score per scenario.

Plan Breadth measured the number of factors considered under the METT-TC structure (i.e., terrain, likely enemy activity, mission/sensor requirements, troops, time constraints, and civilian) with bonus points for considering 2/3 factors in every category. It is reported as percentage of maximum breadth score per scenario.

Sophistication indicates how many advanced features of the SMI system were used in creating plans for unmanned assets. Plan sophistication was rated on a four point scale:

- Zero points were given if the plan was not input in the SMI (i.e., only verbal or on paper).
- One point was awarded for placing a non-mobile asset or “drawing” a straight path from point A to point B.
- Two points were given if multiple waypoints were placed in specific locations taking into account roads, hills, NAIs, or other terrain features.
- Three points were given for using advanced SMI features of plan parameters, point parameters, or actions (i.e. stealth mode, altitude maximum, or take an image) when it was not required.
- Four points were awarded for using advanced features when required to satisfy the OPORD requirements.

Plan sophistication is reported as the percentage of optimal per scenario given the individual participant’s asset choices (e.g., an immobile asset could only earn 1 point whereas using a mobile asset could earn 4 points if advanced features were required). If a participant used 1 immobile asset and 2 mobile assets and no advanced features were required, the maximum possible sophistication was  $1 + 3 + 3 = 7$ . Plan sophistication is only reported for the far transfer scenarios in which all participants used the SMI.

Route is a count of path considerations such as hills, roads, orbits (on aerial vehicles), obstacles, stealth, smoke, altitude, and speed in route planning. There was no optimum or maximum route consideration so it is reported as the average route score per NAI in a scenario.

All the metrics allow comparison across scenarios with different numbers of NAIs.

## RESULTS

All 32 participants had experience with the SMI interface through participation in a previous experiment. Mean age was 25.4,  $SD = 6.5$ . Seventy eight percent were male and 63% were actively pursuing a college degree or had already obtained an undergraduate degree. The Tower of Hanoi (TOH), number of disks (3, 4, or 5) successfully achieved in 10 minutes, was used as an indicator of aptitude. The higher the number of rings achieved, the better the problem solving ability. The mean number of TOH disks successfully achieved was 3.91 with  $SD = 0.64$ .

Participants were randomly assigned to conditions. There were no significant correlations between any of the four individual variables (i.e., age, gender, college degree, TOH) or between the individual variables and the four training conditions. Although, the correlation between TOH and condition did not reach significance, there were no 5-disk TOH level participants in the No Worked Examples condition. TOH was significantly correlated with Total Asset score ( $r = 0.37, p = 0.035$ ) and with the FT2 Asset score ( $r = 0.38, p = 0.031$ ). TOH was positively correlated with Total Sophistication score ( $r = 0.37, p = 0.036$ ) and with FT2 Sophistication score ( $r = 0.34, p = 0.056$ ). TOH was significantly negatively correlated with the following measures: Breadth score for FT1,  $r = -0.34, p = 0.054$  and Breadth score for FT2,  $r = -0.36, p = 0.04$ . There was no correlation with any of the Route scores. Since there were some correlations with TOH and performance, the trials in which there were correlations with TOH were analyzed both with and without the TOH as a covariate since both aptitude as measured by TOH score and experimental training conditions may contribute to performance scores.

Two-way analysis of variance was conducted for each of the four performance measures (Assets appropriately used, Breadth of options considered, Sophistication of plan, and Route considerations) for the first timed training trial (Timed1), the second timed training trial (NearTransfer), and the two far transfer trials (FT1 and FT2). A  $p$  value of .10 was used as the metric of statistical significance.

### First Timed Training

For the first timed training exercise (Timed1), two-way analyses of variance indicated that for Asset score there was a significant main effect of aid,  $F(1,28) = 7.00, p = 0.013$ , and a significant main effect of media,  $F(1,28) = 7.00, p = 0.013$ . Asset scores were better without

worked examples ( $M = 79.16\%$  of optimal,  $SD = 23.96$ ) than with worked examples ( $M = 58.33\%$  of optimal,  $SD = 25.84$ ). Asset scores were better when plans were made on paper ( $M = 79.18\%$  of optimal,  $SD = 23.96$ ) than when they were made using the SMI ( $M = 58.33\%$  of optimal,  $SD = 25.84$ ).

There was no difference between the aid levels on Breadth, or Route scores. There was no difference between the media levels for Breadth scores. Interestingly, while there was no main effect for aid or training media, there was a significant two-way interaction on Route scores,  $F(1,28) = 5.34, p = 0.029$ ; see Figure 4). No Worked Examples was more effective for paper-based training and Worked Examples were better when used with the SMI. This, although not typically significant except for route scores, was a common pattern throughout the experiment.

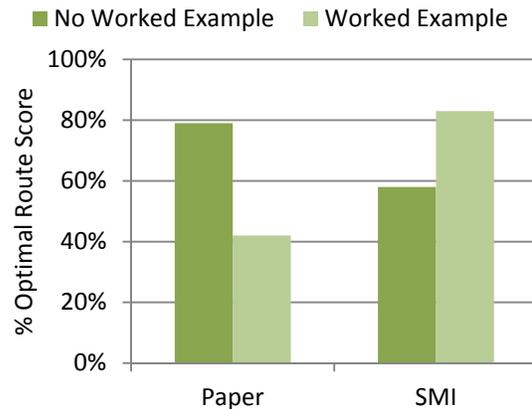


Figure 4. Two-way interaction of aid by media for Route score on the Timed1 scenario

### Near Transfer

The second timed training exercise (NearTransfer) was a conceptual near transfer trial, since it was more difficult than Timed1 and worked examples were not available during the trial. The two-way analyses of variance indicated that there was a non-significant benefit for no worked examples and a significant main effect of media,  $F(1,28) = 5.73, p = 0.026$ . Asset scores were significantly higher with paper than with the SMI (see Table 1).

There was no difference between the aid levels or the media levels for Breadth or Route.

**Table 1. Descriptive statistics for Asset score on the near transfer scenario**

| Worked Example    | Paper vs. SMI | Mean | SD  | N  |
|-------------------|---------------|------|-----|----|
| No Worked Example | Paper         | .75  | .24 | 8  |
|                   | SMI           | .46  | .25 | 8  |
|                   | Total         | .60  | .28 | 16 |
| Worked Example    | Paper         | .58  | .15 | 8  |
|                   | SMI           | .42  | .39 | 8  |
|                   | Total         | .50  | .30 | 16 |
| Total             | Paper         | .67  | .21 | 16 |
|                   | SMI           | .44  | .32 | 16 |
|                   | Total         | .55  | .29 | 32 |

### Far Transfer Exercises

The far transfer exercises were conducted one week after the training session. The first far transfer exercise (FT1) evaluated the transfer from paper to the SMI. For FT1, two-way analyses of variance indicated a non-significant benefit for SMI on Asset assignment, and this held whether training used worked examples or not (see Table 2). That is, there was no significant benefit or loss to training on paper versus the SMI with respect to transfer performance on the SMI.

There was no difference between the aid levels or the media levels for Breadth or Route. The differences favoring worked examples and SMI were not significant on Sophistication scores.

**Table 2. Descriptive statistics for Asset score on FT1**

| Worked Example    | Paper vs. SMI | Mean | SD  | N  |
|-------------------|---------------|------|-----|----|
| No Worked Example | Paper         | .25  | .24 | 8  |
|                   | SMI           | .46  | .35 | 8  |
|                   | Total         | .35  | .31 | 16 |
| Worked Example    | Paper         | .37  | .33 | 8  |
|                   | SMI           | .42  | .15 | 8  |
|                   | Total         | .40  | .25 | 16 |
| Total             | Paper         | .31  | .28 | 16 |
|                   | SMI           | .44  | .26 | 16 |
|                   | Total         | .37  | .28 | 32 |

The second far transfer exercise (FT2) was conducted on the SMI for all training groups and evaluated a conceptual transfer to a more complex task with new problem events. The two-way analyses of variance indicated that for Asset selection, there was a main effect of aid,  $F(1,28) = 10.23, p = 0.003$  and a main effect of media,  $F(1,28) = 4.79, p = 0.037$ . Asset

scores were higher when participants trained without worked examples and when they trained on the SMI (see Table 3). When an analysis of covariance was conducted with TOH as the covariate, these effects were still significant for aid,  $F(1,27) = 6.87, p = 0.014$ , and for media,  $F(1,27) = 3.98, p = 0.056$ .

The differences favoring No Worked Examples and SMI on Sophistication score were not significant. There was no difference between any condition for Breadth or Route. However, a subsequent analysis of covariance conducted with TOH as the covariate, due to the significant negative correlation between TOH and Breadth, revealed that the benefit for No Worked Example was significant on Breadth score,  $F(1-27) = 5.57, p = 0.026$ .

**Table 3. Descriptive statistics for Asset on FT2**

| Worked Example    | Paper vs. SMI | Mean | SD  | N  |
|-------------------|---------------|------|-----|----|
| No Worked Example | Paper         | .88  | .23 | 8  |
|                   | SMI           | 1.00 | .00 | 8  |
|                   | Total         | .94  | .17 | 16 |
| Worked Example    | Paper         | .50  | .40 | 8  |
|                   | SMI           | .78  | .25 | 8  |
|                   | Total         | .64  | .35 | 16 |
| Total             | Paper         | .69  | .37 | 16 |
|                   | SMI           | .89  | .20 | 16 |
|                   | Total         | .79  | .31 | 32 |

### Total score over all exercises.

Two-way analyses of variance were conducted for the total scores over all six exercises. Results indicated a significant main effect of aid,  $F(1,28) = 6.51, p = 0.016$  for Asset assignment. Asset scores were significantly higher without worked examples ( $M = 68.02\%$  of optimal,  $SD = 9.38$ ) than they were with worked examples ( $M = 57.35\%$  of optimal,  $SD = 13.51$ ). This was still significant,  $F(1,27) = 3.86, p = 0.060$  when an analysis of covariance was conducted with TOH as the covariate.

There was no overall difference between the worked example conditions or the paper/SMI training conditions on Breadth. For Route scores, there were no differences between the worked example conditions. The SMI training was not significantly different than paper with respect to Route. The interaction was significant,  $F(1,28) = 2.19, p = 0.085$  indicating that Worked Examples were more beneficial with the SMI environment and No Worked Examples was more

beneficial with Paper. Again, this pattern, although not always significant, was common across many trials and measures (e.g., FT2-route, FT1-route, NearTransfer-route, Timed1-route).

### **Performance Time**

Performance Time was relatively similar for the training and transfer trials. The two-way analysis of variance revealed no significant differences in trial performance time for Timed1, NearTransfer, or FT2. For FT1, there was a main effect of media,  $F(3, 28) = 3.17, p = 0.086$ . Performance time for the SMI trained group ( $M = 887.19$  seconds,  $SD = 300.70$ ) was significantly faster than the paper group ( $M = 1090.37$  seconds,  $SD = 355.75$ ), as expected.

## **DISCUSSION**

There are several themes that can be gleaned from the results. First we discuss the implications for TOH as a test of aptitude. Second we discuss the usefulness of worked examples and learning media.

A computerized version of the TOH puzzle was used to assess general problem solving ability. The TOH has been used as a test of planning and problem solving ability under low training conditions (O'Donnell et al 2005). For the Tower of Hanoi problem, the highest number of disks solved in a 10 minute period was used as a measure of problem solving ability. Participants started with 3 disks and repeated it until they could solve it in the minimum number of moves. Likewise, they progressed to 4 and 5 disks. They stopped when they mastered 5 disks or when 10 minutes elapsed. Participants were classified by the maximum number of disks mastered. In a previous study on exploratory learning with the SMI, there was a significant positive correlation between TOH disks and far transfer score (Carolan, Hutchins, McDermott and Wickens, 2011).

In the current experiment, TOH was positively correlated with Asset Score (Total and FT2) and with Sophistication (Total and FT2). The higher the TOH score, the better the asset choice and the greater number of advanced features were used in the SMI. This provides another piece of evidence in support of using the TOH as a measure of problem solving ability. The advantage is that TOH is quick and easy to administer.

TOH was found to be negatively correlated with Breadth scores in the transfer exercises. Taken together, this shows that those with higher TOH scores

were performing well (as evidenced by the Asset score) but were not considering as many factors as other participants. The smaller number of considerations reported did not have a detrimental impact on performance. Breadth scores may have been lower because the participants had adopted strategies in which they focused on key diagnostic cues and therefore were able to make asset choices without considering a large list of factors. For example, if an NAI was suspected of having improvised explosive devices (IEDs) hidden by the enemy, this could immediately limit the number of appropriate assets to the SUGV and the XUV assets. A discerning participant could recognize that only the XUV could do target acquisition and therefore was the appropriate choice. Not all choices could be made by only considering two factors but given vehicles and situations, certain factors were more diagnostic than others.

Asset scores for the near transfer scenario were higher for those using paper than those using the SMI. However, for the far transfer scenarios, there was no difference in Asset scores. Participants in the paper conditions were slower on the first far transfer scenario which was the first scenario in which they planned using the SMI. Even though those in the paper condition took longer to complete the first far transfer scenario, their Asset scores did not suffer. By the second transfer scenario, their planning times were no different than those who had practiced on the SMI. It is important to note that all participants had previous experience inputting routes on the SMI. It would be interesting to evaluate how well these findings are replicated in a similar experiment with completely novice SMI users. The far transfer performance suggests that using a paper-based planning approach is an effective strategy when the operational system is not available.

The use of worked examples did not seem to provide a training advantage in this research study. In fact, those without worked examples made better asset choices. Those without worked examples had better Asset scores overall, on the Timed1 scenario, and most importantly, on the second (adaptive) far transfer scenario. This is not inconsistent with previous research that suggests the use of worked examples may be limited for more complex problems requiring more adaptive solutions. The worked examples helped learners to get an understanding of how to approach the problem earlier in the process but did not seem to provide a longer term cost or benefit.

On a related note, there was an interaction observed between media and aid. Those without worked examples did better when trained on paper and those with worked examples tended to do better when trained on the SMI. Granted, this was only significant for route but it was a common non-significant result throughout the data. Worked examples may be more appropriate when trainees are to learn a complex cognitive task in conjunction with digital system procedures. The worked example may be more useful when they are attempting both types of tasks at once.

Those without worked examples also had significantly higher Breadth scores on the second far transfer scenario. These participants continued to consider a larger variety of factors in choosing between unmanned vehicles. The implication is that if consideration of a wide variety of factors is a critical part of the transfer task, then it may not be advantageous to use worked examples during training.

The link between Breadth and Asset score is interesting. Those with high TOH scores tended to have high Asset scores and low Breadth scores in transfer. However, those without worked examples tended to have higher Asset scores and higher Breadth scores than those with worked examples. High Asset scores could be associated with either high or low Breadth scores. This could be due to different strategies used to arrive at appropriate asset decisions. A participant may focus on a few diagnostic cues or they may consider a wide range of cues but both can be successful in terms of asset selection. This is another indication of the richness of the problem domain.

## **CONCLUSION**

In summary, we found limited evidence of the usefulness of worked examples in a complex planning task coupled with procedural digital tasks. Participants who did *not* have the worked example scored better on asset selection in the transfer scenarios. Asset selection can be likened to “goodness of plan content.” Having worked examples may be advantageous in situations in which it is important to consider multiple factors when transferring to a real world task. We also found that the TOH was useful as a measure of planning aptitude and should be considered as a quick and easy method of assessing planning and problem-solving aptitude.

## **ACKNOWLEDGMENTS**

This work is supported by the US Army Research Institute under Contract #: W91WAW-09-C-0081 titled Understanding the Impact of Training on Performance. The view, opinions, and/or findings contained in this paper are those of the author (s) and should not be construed as an official Department of the Army position, policy, or decision.

## **REFERENCES**

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70, 181–214.
- Carolan, T. F., Hutchins, S., McDermott, P.L., & Wickens, C.D. (2011). Understanding the Impact of Training on Performance: Phase II Option Year I Interim Report. Report submitted to the US Army Research Institute, Contract #: W91WAW-09-C-0081
- Hutchins, S.D., Wickens, C.D., Carolan, T.F., Cumming, J. (2012 submitted). Error Prevention Training Support: Meta-Analyses of Transfer. *Human Factors Journal*, submitted 2012.
- Kalyuga, S. (2007). Rapid dynamic assessment of expertise to improve the efficiency of adaptive E-learning. *ETR&D*, 53(3), 83–93.
- O'Donnell RD, Moise S, Schmidt RM. (2005). Generating performance test batteries relevant to specific operational tasks. *Aviation, Space, and Environmental Medicine*, 76 (7 Suppl), C24-30.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, 23, 90-108.
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review* 22(2), 155-174.