# Adaptive Training Strategies:
# Considering the Decision Process as well as the Outcome

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

## ABSTRACT

Improving the efficiency of training complex cognitive tasks necessitates controlling cognitive load during learning. Adaptive training manages the cognitive load by adjusting the training schedule and difficulty to match immediate performance. This experiment is innovative in both the complexity of the experimental task and the adaptation scheme. The complex cognitive planning task required participants to choose and route unmanned vehicles to reconnoiter geographic areas. Each scenario had multiple suitable solutions and there were many irrelevant, relevant, and key factors that could be considered with regards to the vehicles, terrain, and mission. Adaptation was based on both performance and decision process (i.e., number of factors considered). We compared two types of adaptation to a typical graduated-difficulty condition and a constant control. When all training conditions were combined, there was a significant correlation between key factors and performance. The more key factors considered, the better the performance. Adapting by decision process (as opposed to adapting by performance alone) had a marginal effect on performance. For those that started with low difficulty, those that adapted by decision process made better plans than those that adapted by performance alone. This study provides evidence on the usefulness of adapting by decision process and warrants further research on the topic.

## 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.

**Mark Gronowski** is an Associate Cognitive Scientist at Alion Science and Technology. He has a B.A in Psychology from the University of Colorado in Boulder, CO and has worked on the design and conduct of multiple empirical studies of training effectiveness.

**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.

# Adaptive Training Strategies:
# Considering the Decision Process as well as the Outcome

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

## INTRODUCTION

Training complex cognitive tasks requires careful management of trainee cognitive load. Task complexity, such as monitoring ongoing portions of the task while simultaneously gathering information for a different portion of the task, can increase the cognitive effort required and subsequently decrease learning. Cognitive load theory provides an important framework for managing learner effort and workload (i.e., Sweller, 1988, Paas, Renkl, Sweller, 2003; Van Merrienboer, Kester & Paas, 2006; Paas & van Gog, 2009). Adaptive training works within this framework to promote more effective learning by adjusting the task to best suit the learner's current skill level across training (Wexley & Latham, 1991).

There is a history of support for adaptive training. Landsberg et al. (2012) summarize several methods for *how* to adapt training (aptitude, prior knowledge, and performance) and *what* should be adapted (instructional content or task difficulty). They suggest that adapting based on performance can improve the effectiveness of training as well as its efficiency (i.e. Tennyson and Rothen, 1977). Adapting the difficulty of a training task based on the performance of a learner can effectively tailor training to suit a specific participant's needs; e.g. if a learner is performing well, the task difficulty can be increased (Wexley & Latham, 1991).

This was corroborated in a recent meta-analysis. Wickens, Hutchins, Carolan and Cumming (2012) found that increasing difficulty should be done adaptively in order to improve performance. Wickens et al. found that there was neither a cost nor benefit for increasing difficulty when compared to a control. However, there was a significant benefit of increasing difficulty adaptively as compared to keeping difficulty constant. But while adaptive sequencing was more beneficial that constant difficulty, constant difficulty was more beneficial than fixed difficulty sequencing.

Spain's et al. (2012) review of current trends in adaptive training research found that there is a missing link in the research regarding what, how and when to adapt training. While they expand on macro-adaptive training (training that contains pre-planned decisions on what to adapt and how to adapt it based on prior knowledge, pretests, etc.) they suggest a more promising solution in micro-adaptive training. Micro-adaptive training is training that implements ongoing assessment of student performance to determine how, what, and when to adapt training. Training can incorporate micro-adaptive strategies by sequencing the content according to the performance of the student. An ideal micro-adaptive training method incorporates tailored training techniques that use feedback, prompts, or hints based on response errors and even emotional states.

Not all task types have been addressed with adaptive training. Wickens et al. found only two training experiments that involved complex cognitive tasks (Barret et al., 1977; Salden et al., 2004). Barret et al. used perception-reaction tests to determine the most effective training for concept-formation tasks. Salden et al. used a higher complexity air traffic control task. We addressed this gap by designing an experiment with a complex, non-trivial cognitive task in the domain of Army planning. The task involved making decisions about which unmanned asset to send to specific geographic regions and creating the routes for those vehicles.

The innovation in this research is that task difficulty is not just adapted based on performance, but on the decision process as well. Examining decision process provides additional evidence as to how a trainee deals with the challenges and nuances of the task. Rather than adapting training by using only an end-result performance metric, the decision process itself can indicate whether a trainee should receive more intense remediation or, more importantly, higher advancement. We presumed that considering both performance and decision process would result in a more accurate assessment of their understanding of the task problem and therefore their cognitive

readiness for the current difficulty level. Adapting the training in this manner allowed participants to advance or decline according to their performance but change the magnitude of the adaptation based on their consideration of information.

In terms of performance, based on results of a recent meta-analysis (Wickens et al. 2012) we expected Constant to be superior to Fixed sequencing and we expected Adaptive Up to be superior to Constant sequencing. This paper primarily focuses on the decision process engendered by the two adaptive training strategies. Post-experiment, we classified information considered as key, relevant, and irrelevant to the decisions to be made. We then examined participant strategies to determine if the different training strategies resulted in more diagnostic strategies. In other words, did they consider more key and relevant factors and fewer irrelevant factors? Our driving research questions are focused on adapting based on the decision process.
1. Do people who focused on key factors do better?
2. Did certain training conditions lend themselves to homing in on key factors?
3. Did adaptive training based on decision strategy help performance? Did it impact consideration of scenario factors?

The objective was to research the impact of adaptive training on the ability to adapt to a more complex decision making scenario.

## METHODS

### 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 for conducting reconnaissance on Named Areas of Interest (NAIs) and determined how to employ them (i.e., send them on a route that uses roads, avoids hills, or flies an orbit around the area of interest). 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, participants were given a manual of vehicle characteristics. Participants reviewed the manual for up to ten minutes prior to the first training scenario and used it as a reference guide throughout the experiment. The vehicle capabilities were broken down into the seven main considerations of METT-TC: Mission, Enemy, Terrain, Time, Troops available, and Civilians. The manual was used to understand the advantages and disadvantages of the different assets for different situations. A limited number of vehicles (2-5) were available in each scenario and the participant's task was to determine whether and how to employ them. For a more detailed description, see McDermott, Carolan, and Gronowski (2012) which describes a different experiment that used similar OPORDs and the same manual of vehicle characteristics.

### Experimental Design

The complexity of the training scenarios varied based on the size of the problem. Each level added another NAI, another asset, more key and relevant pieces of information in the OPORD, and terrain obstacles (i.e., hills, forests, lakes, roads, towns). Difficulty level 1 scenarios had one NAI, two assets, an average of 6 key/relevant factors and open flat terrain with few roads. Difficulty level 3 added timing factors. Participants needed to understand the timeline of the mission including current time, mission start time, and when each NAI needed coverage during the mission. Difficulty level 4 had an average of 22 key/relevant factors, extra NAI specific information, terrain considerations that were not listed in the OPORD (they had to be interpreted from the map), and more ambiguity between courses of action.

Practice Order was a between-subjects variable with 4 levels:
- Constant: scenario complexity was always at level 4, the highest level of training complexity;
- Fixed Progressive: scenario complexity increased incrementally from low level complexity to full complexity;
- Adaptive Up: Scenario complexity began low (level 1) and adaptively increased or decreased depending on performance;

- Adaptive Down: Scenario complexity began high (level 4) and adaptively increased or decreased depending on performance

Participants were matched to conditions using a pre-test of problem solving ability (e.g., Tower of Hanoi puzzle).

**Participants**

Sixty participants completed the experiment. This included 41 men and 19 women, ranging in age from 18 to 40 years ($M = 27.15$, $SD = 5.51$). Participants were recruited from Craigslist and had to achieve the 30th percentile of a normal adult population of the Ravens' Standard Progressive Matrices (a score of 19 out of 28) in order to qualify for the experiment. Roughly 10% of participants did not qualify. Participants were compensated $25 for the first session (1 hour), $50 for the second session (4 hours), and $100 for the third session (2 hours) for a total of $175. The payment schedule was structured to encourage participants to return for the third session.

**Apparatus and Stimuli**

Two assessments of problem solving ability were utilized: Tower of Hanoi and Raven's Standard Progressive Matrices. A computerized version of the Tower of Hanoi (TOH) puzzle was used to assess memory for procedures and general problem solving ability. Participants began with 3 rings and repeated it until they could solve it in the optimum number of moves. Likewise, they progressed to 4 and 5 rings. They stopped when they mastered 5 rings or when 10 minutes elapsed. Participants were classified by the maximum number of rings mastered. The Raven's Standard Progressive Matrices (Raven, 2000) offers insight about one's capacity to observe, solve problems, and learn. Problems are presented as a matrix of patterns or sequence of diagrammatic puzzles with one piece missing. Participants complete the pattern or sequence by choosing the correct missing piece from a list.

Fifteen paper OPORDs were developed in collaboration with Army subject matter experts. The OPORDs contained information about the: Situation (weather, expected enemy action, adjacent friendly activities), Mission (assets available, goals, deadlines), Execution (mission timeline, data collection needs, coordinating instruction, enemy likelihood in different areas, rules of engagement), and Service Support (logistics, special weapons).

An 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. Assets were ranked as preferred, acceptable, and unsuitable for different situations. A half-page outline described the four major steps in creating a plan.

Demographic data regarding age, gender, education, and computer game experience was collected. A post-experiment questionnaire solicited strategies and reactions to the training method.

**Procedure**

*First Session: Screening*

Participants worked on the computerized Tower of Hanoi puzzle for ten minutes and the Raven's Standard Progressive Matrices for up to 47 minutes.

*Second Session: Training*

Instructor interaction with participants was minimized in order to maintain consistency between participants. The instructor described the general task, the Asset Manual, and the outline of basic steps. Participants took up to 10 minutes to 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, although some answers were clearly better than others. The focus was on what information was considered when making their decisions. Participants were told that the instructor would ask questions about the information they considered when creating the plan. In addition to choosing an asset or assets for each NAI, participants drew routes to indicate where the asset should travel. For air assets, they were instructed to draw orbits. For all assets, participants were instructed to consider whether they want to use or avoid hills and roads.

Feedback was corrective for all conditions. Scenario order varied by training condition. For the Adaptive conditions, the first scenario was prescribed but subsequent scenarios depended on performance. The Adaptive Up condition began with complexity level 1 and the Adaptive Down training condition began with complexity level 4. The rules for determining subsequent scenarios were as follows:

- If all NAIs had suitable assets, the participant passed and moved up 1 complexity level. If the participant passed and considered extra factors (150% of minimum relevant factors) they moved up 2 complexity levels
- If one or more NAIs had an unsuitable asset, the participant failed and moved down 1 level of complexity. If the participant failed and considered less than 50% of the minimum relevant factors, they moved down 2 complexity levels.

There were two available scenarios for complexity levels 1, 2, and 3. There were four scenarios for complexity level 4. Each scenario was only attempted once. If a participant had already completed their "next" scenario, they would complete the closest scenario in terms of complexity. For example, if a participant failed complexity level 3 and had to move down two levels but the complexity level 1 scenarios were complete, the participant would attempt complexity level 2 instead. The Adaptive Down condition had to complete a minimum of three scenarios; they could not pass the first or second scenario and be done. All participants ended with the near transfer scenario, 4D. It was near transfer because it included a new asset that participants had not previously encountered. The new asset was a type of UGS but instead of being placed by Soldiers it could be launched from a distance. This had implications for its range and need for logistics support. Participants were given a 20 minute break after one hour and 40 minutes. Some participants did not take that long and thus did not take a break during the session.

*Third Session: Transfer*

The third session occurred one week after the first session and lasted two hours. Participants took up to 5 minutes to look over the manual before starting scenarios. Participants had access to the asset manual and lined paper for taking notes. Participants completed the following scenarios:

*Scenario 5*. This near transfer scenario was challenging because there was a large number of factors present, the urban terrain was new, and two assets were suitable for NAI 2, although one asset was highly preferred over the other. The time limit was 16 minutes.

*Scenario 5 Replan*. This was a far transfer test of participants' ability to rapidly replan given new information. After participants described their solution for scenario 5, participants received a weather update of impending rain. They had 2 minutes to decide if and how to change their asset allocation. The correct solution was to realize that the Goldeneye is unsuitable for rain and to use the REMBASS instead.

*Scenario 6.* This near transfer scenario was challenging due to the large number of factors present and the fact that there was only one factor against using the Stryker in each of the NAIs which meant more ambiguity in the decision; all previous scenarios had 2 or more factors against using a vehicle at an NAI. Time limit was 16 minutes.

*Scenario 6 Replan.* This far transfer scenario tested ability to replan when more than one factor changed in a scenario. The commander identified a new NAI that needed to be covered. The new NAI had terrain, mission, and timing constraints. To further complicate things, the participant was given one new asset to create a plan. This was in addition to the 5 assets available for the initial plan. The new asset was unsuitable for the new NAI so participants had to reallocate one of their initial assets to cover the new NAI and then use other assets to cover the gap left by reallocating an asset. Participants had 6 minutes to replan and draw new routes on the paper map.

*Scenario 7*. This far transfer scenario tested ability to choose between equally suitable assets. Participants were asked, "If you could have one additional piece of information about each NAI, what would it be?" Experimenters rated how "diagnostic" that piece of information was – how well it helped differentiate between the two suitable assets. For example, the SUGV and the XUV were both suitable for NAI 3. Knowing whether motion imagery was required would be highly diagnostic because needing motion imagery would eliminate the XUV from consideration. The time limit was 14 minutes.

Participants were given a warning when they had 5 minutes remaining and 1 minute remaining. After each scenario, participants articulated the factors they considered in making their decision.

**Dependent Measures**

There was not one right answer or plan for each scenario. Therefore, two dependent measures were used to make quantitative comparisons between plans: asset score and plan breadth. Asset score is a measure of plan correctness. An NAI with an acceptable asset that contained no unsuitable characteristics was worth 2 points. An NAI with an unsuitable asset scored zero points. Asset score is a percentage of optimal, which allows comparisons across scenarios with different numbers of NAIs.

Scenarios contained information, or factors, about the mission that mapped to asset capabilities. For each NAI in each scenario, the full set of factors was classified as key, relevant, and irrelevant according to the following definitions:
- A key factor is a factor that is unsuitable for an asset for an NAI. It is a piece of information that enables one to rule out" an asset for a particular NAI
- A relevant factor is a factor that does not eliminate an asset but makes it preferred over another. When two vehicles are both suitable for an NAI, a relevant factor will help make the decision on which one is better employed at an NAI. These factors are typically present in more difficult scenarios.
- An irrelevant factor is a factor that was not present in the scenario.

The list of key, relevant, and irrelevant factors was different for each NAI in each scenario. The dependent measure is the count of key, relevant, and irrelevant factors considered per scenario.

Training Duration was the total amount of time spent during training and practice in Session 2, starting with the training overview and ending with the completion of the last training scenario.

**RESULTS**

Correlations using Pearson's r were run between training condition and asset score, key factors, relevant factors, and irrelevant factors. An alpha level of .05 was used.

**Did those who focused on key factors do better? Did they consider more factors in general or did they discriminate and focus on factors that differentiated asset usefulness for the situation?**

Overall, the number of key factors considered was significantly positively correlated with asset score ($r = 0.425$, $p < 0.001$). Neither the number of relevant factors, nor the number of irrelevant factors was correlated with asset score. The number of key and relevant factors considered was correlated ($r = 0.670$, $p < 0.001$) and the number of relevant and irrelevant factors considered was correlated ($r = 0.405$, $p < 0.01$).

This correlation pattern was consistent for the three primary transfer scenarios (Table 1).
- Key factors and asset score was positively correlated for scenarios 5, 6, and 7.
- Irrelevant factors and asset score were significantly negatively correlated in scenario 7. Overall and individual scenarios followed this trend, though not to significance.
- The number of key and relevant factors was positively correlated for scenarios 5, 6, and 7.
- The number of relevant factors was correlated with asset score in scenario 6.
- The number of irrelevant and relevant factors considered was positively correlated for scenarios 5, 6, and 7.

*Best and Worst performers*
The correlation results do not provide a clear indication if the findings are due primarily to the structure of the scenarios (e.g., number of key and relevant factors required for asset assignment) or if better performers considered more key and relevant factors, and fewer irrelevant factors, than poorer performers. Therefore, we split the participants into performance groups to explore the overall differences in factor consideration between good and poor performers and changes in factor consideration over the three scenarios.

The 60 Participants were split into four subgroups based on their overall asset performance. Where the asset scores were the same above and below a quartile, participants with the same asset score were included in the higher group.

The lower limits for the Top quarter was .85 ($n=16$), 2nd quarter was .77 ($n=15$), 3rd quarter was .62 ($n=17$), and 4th quarter was .61 ($n=12$).

**Table 1. Correlation data for asset score and factors**

| Correlation | *r* | *P* |
|---|---|---|
| Asset score and key factors | | |
| Scenario 5 | 0.319 | < 0.05 |
| Scenario 6 | 0.575 | < 0.001 |
| Scenario 7 | 0.352 | < 0.01 |
| Key factors and relevant factors | | |
| Scenario 5 | 0.594 | < 0.001 |
| Scenario 6 | 0.595 | < 0.001 |
| Scenario 7 | 0.577 | < 0.001 |
| Asset score and relevant factors | | |
| Scenario 6 | 0.259 | < 0.05 |
| Relevant factors and irrelevant factors | | |
| Scenario 5 | 0.320 | < 0.05 |
| Scenario 6 | 0.394 | < 0.01 |
| Scenario 7 | 0.301 | < 0.05 |
| Asset score and irrelevant factors | | |
| Scenario 7 | -0.269 | < .05 |

An ANOVA was used to evaluate differences between the four asset performance groups on key, relevant, and irrelevant factor consideration (Table 2). The differences in mean asset performance were of course significant ($p < 0.0001$) overall and between all quartiles. For the overall factors considered, there was a main effect for asset quartile for key factors, $F(3,56) = 4.5$, $p < 0.01$, and relevant factors, $F(3,56) = 2.7$, $p = 0.05$. There was no main effect for irrelevant factors. Unless noted, the p values for any differences between quartiles are at least 0.05. The top quarter considered significantly more key factors than either the 3rd or 4th quarters and the 2nd quarter more than the 4th quarter. For relevant factors, the top quarter considered significantly more than the 3rd and marginally more than the 2nd quarter ($p = 0.09$). As Table 3 shows, the consideration of key factors was relatively linear with respect to quartiles. However, consideration of relevant factors clearly separates the top performers from the rest of the performance groups.

**Table 2. Total Factors (Scenarios 5, 6, 7)**

| Quartile | Mean Key Factors (*SD*) | Mean Relevant Factors (*SD*) | Mean Irrelevant Factors (*SD*) |
|---|---|---|---|
| Top | 25.38 (4.08) | 21.00 (8.70) | 20.75 (10.83) |
| 2nd | 24.07 (5.75) | 16.13 (9.02) | 18.80 (9.51) |
| 3rd | 21.18 (4.89) | 13.35 (5.92) | 15.35 (7.96) |
| Bottom | 19.00 (4.33) | 16.58 (7.06) | 25.42 (17.94) |
| Total | 22.58 (5.28) | 16.73 (8.11) | 19.67 (11.88) |

To determine if factor consideration improved over the course of the transfer scenarios, analyses of factor consideration were conducted for each of the three transfer scenarios. On Scenario 5, there was a significant main effect of quartile group on key factors, $F(3,56) = 2.74$, $p = 0.05$. The top group differed significantly from the fourth quarter, and marginally from the third quarter ($p = 0.06$). The means were 6.13 for the top quarter, 5.06 for the third quarter, and 4.42 key factors for the 4th quarter. There were no main effects for relevant factors or irrelevant factors.

On Scenario 6 (Table 3), the mean differences between performance groups on key factors was highly significant, $F(3,56) = 9.55$, $p < 0.0001$. The top quarter considered significantly more fey factors than the 3rd and 4th quartiles, and the 2nd group mean was also significantly higher than the 3rd and 4th quartiles. The mean difference on relevant factors considered was also significant, $F(3,56) = 2.85$, $p < 0.05$. The top quartile mean for relevant factors considered was significantly more than the 3rd and marginally more than the 4th ($p = 0.08$) quartiles, and the 2nd quartile mean was also significantly more than the 3rd quartile. The number of irrelevant factors considered was not

significant. So, it appears that by Scenario 6, the top two quartiles were considering more key factors and starting to consider more relevant factors.

**Table 3. Scenario 6 Factors**

| Quartile | Mean Key Factors (*SD*) | Mean Relevant Factors (*SD*) | Mean Irrelevant Factors (*SD*) |
|---|---|---|---|
| Top | 10.25 (1.53) | 5.19 (2.14) | 7.06 (4.60) |
| 2nd | 9.33 (1.84) | 4.73 (2.74) | 6.87 (3.52) |
| 3rd | 7.71 (2.39) | 3.29 (1.40) | 4.88 (3.33) |
| Bottom | 6.58 (2.11) | 3.83 (1.53) | 7.00 (6.33) |
| Total | 8.57 (2.39) | 4.27 (2.12) | 6.38 (4.44) |

By Scenario 7 (Table 4), the main effects for performance group were significant for key factors, $F(3,56) = 2.99$, $p < 0.05$, relevant factors, $F(3,56) = 3.55$, $p < 0.05$, and irrelevant factors, $F(3,56) = 3.34$, $p < 0.05$. While the top quartile considered more key factors than the 3rd and 4th quartiles, it did not differ significantly from the 2nd group. The top group considered significantly more relevant factors than either the 2nd, 3rd or 4th quartiles. And the bottom group considered more irrelevant factors than the top, 2nd or 3rd quartiles.

**Table 4. Scenario 7 Factors**

| Quartile | Mean Key Factors (*SD*) | Mean Relevant Factors (*SD*) | Mean Irrelevant Factors (*SD*) |
|---|---|---|---|
| Top | 7.69 (1.30) | 7.44 (2.78) | 4.31 (2.39) |
| 2nd | 6.87 (2.67) | 5.40 (3.46) | 4.27 (4.17) |
| 3rd | 6.18 (1.38) | 4.47 (2.10) | 3.41 (2.94) |
| Bottom | 5.83 (1.64) | 5.25 (2.22) | 7.42 (4.42) |
| Total | 6.68 (1.91) | 5.65 (2.87) | 4.67 (3.69) |

**What is the impact of training condition on performance and on the consideration of factors?**

*Performance*
The overall differences across the four training conditions on Asset score did not reach significance.

There was a significant main effect of training duration, $F(3,53) = 11.9$, $p < 0.001$. Comparisons between conditions found that Constant difficulty was completed in significantly less time than either Adaptive Up or Adaptive Down. Fixed difficulty was also significantly faster than either Adaptive Up or Adaptive Down. The data for training times is listed in Table 5.

**Table 5. Training Times**

| | N | Mean Time (minutes) | SD |
|---|---|---|---|
| Constant | 15 | 99.751 | 17.3187 |
| Fixed | 15 | 113.671 | 23.1339 |
| Adaptive Up | 14 | 156.584 | 49.4547 |
| Adaptive Down | 13 | 165.269 | 41.7838 |
| Total | 57 | 132.316 | 43.8993 |

*Factor consideration*
One way ANOVAs were conducted to investigate potential differences between training conditions for asset score and for key, relevant and irrelevant factors.

For the overall analysis, the number of irrelevant factors considered indicated a significant difference across the four training groups, $F(3,56) = 3.07$, $p < .05$. Post-hoc comparisons indicated that the Fixed training group considered

significantly more irrelevant factors than either the Constant ($p < 0.05$) or the Adaptive Down groups ($p < 0.05$). Notably the Fixed group also considered significantly more relevant factors than the Constant group ($p< 0.05$) and marginally more than Adaptive Down ($p = 0.06$). Total factors considered was therefore also significant, $F(3,56) = 2.97$, $p < 0.05$.

This pattern was also consistent for the three main transfer scenarios. Fixed scenario difficulty increases in training resulted in the consideration of significantly more relevant and irrelevant factors than either Constant or Adaptive Down scenario difficulty in Scenario 5. This was also found in Scenarios 6 and 7 although the difference was marginally ($p < 0.10$) significant.

**Did those who adapted based on decision process do better or have different strategies?**

For those 29 participants in the adaptive conditions, the data was examined to determine whether a participant ever had a "bonus" adaptation by considering extra factors or too few factors. This bonus adaptation meant that they either increased or decreased difficulty by two levels instead of one level. Participants who did were classified as "adapting by decision process." Participants who did not were classified as "adapting by performance alone." Nineteen participants adapted by decision process and ten participants adapted by performance alone.

*Performance*
Although there were no statistically significant differences, the mean Asset score, key factors considered, and relevant factors considered was higher for participants that adapted by decision process than participants that adapted based on performance alone. This trend was persistent for the overall means as well as the means in every transfer scenario. Another interesting trend is visible when training condition is crossed with adaptation type (see Figure 1). The mean asset scores for Adaptive Down by decision process, Adaptive Down by performance alone, and Adaptive Up by performance alone were very similar, between .67 and .69. However, at .81 the Adaptive Up by decision process group appears to have a higher mean asset score. This was not a planned comparison and the number of participants in each group was small, between 4 and 10. Adapted by decision process versus performance alone was contrasted within Adapted Up alone and within Adapted Down alone. There were no significant results for Adaptive Down; but for Adaptive Up the overall asset difference was marginally significant $F(1,13) = 3.47$, $p = 0.085$.
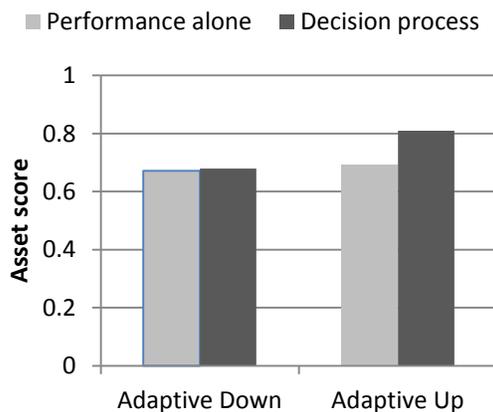


**Figure 1. Asset score by training condition and adaptation type**

In general, those who adapted by decision process followed a pattern relating to their starting point. Those in Adaptive Up tended to adapt up two levels for considering extra factors and those in the Adaptive Down condition tended to adapt down two levels for considering too few factors.

*Correlation between Asset Score and Key Factors*
There was a significant positive correlation between asset score and key factors for the twenty participants who adapted by decision process ($r = 0.612$, $p = 0.004$) as well as the nine participants who adapted by performance alone ($r = 0.714$, $p = 0.031$). Looking at the training condition crossed by adaptation type revealed that the following groups had a significant positive correlation between asset score and key factors: Adaptive Down by decision process ($r = 0.658$, $p = 0.028$, $n = 11$), Adaptive Up by decision process ($r = 0.761$, $p = 0.017$, $n = 9$), and Adaptive Up by performance alone ($r = 0.830$, $p = .041$, $n = 6$). There was no significant correlation in the Adaptive Down by performance alone group ($n = 3$).

**DISCUSSION**

The data provide insight on the relationship between information used and plan quality. Key factors are those that eliminate suitability of an asset for a particular assignment. Better plans (i.e., higher asset scores) were associated

with the consideration of a higher number of key factors. This was true overall and for each primary transfer scenario. Relevant factors are pieces of information in the OPORD that could be used to discern whether an asset was acceptable or suitable for an NAI. They were useful in the decision making process but not as discriminating as key factors which could be used to "rule out" an asset for an NAI. The number of relevant factors and key factors was related, indicating that participants considered relevant factors alongside key factors. The number of relevant factors and irrelevant factors was related, indicating that participants considered relevant and irrelevant factors together. The number of relevant factors considered was only correlated with asset score in one scenario. Thus, the number of key factors considered is a better indication of information discrimination.

It appeared that the participants were still refining their decision process during the transfer scenarios. By the last transfer scenario, there was a negative correlation between irrelevant factors and asset score. Those with better asset scores tended to consider fewer irrelevant factors. This pattern of correlations shows that by the end of testing, a successful strategy involved focusing on key factors and not considering irrelevant factors. This shows that participants who developed high quality plans were discriminating in regards to the type of information that they attended to. They were not just considering a high number of factors; they focused more on high-value information that allowed them to differentiate between asset suitability for a specific NAI.

A significant difference was found between in the Adaptive Up group that adapted by decision process and those that adapted by performance alone. Within the Adaptive Up condition, better performance was seen in those that adapted by decision process. It is not clear if that group performed better because they tended to consider more factors or if performance increased because the adaptation strategy took into account their decision process.

Future research is warranted on the technique of adapting task difficulty based on decision performance, especially in the context of complex cognitive tasks. In this experiment, the determination of whether to adapt by decision process was based on the number of factors considered. We did not consider *which* factors they considered and whether those factors were key, relevant, or irrelevant. In future experiments, we would be more specific. For example, increase difficulty if a high number of key factors was considered or decrease difficulty if a high number of irrelevant factors was considered.

Lastly, we examined differences between the four original training conditions: Fixed, Constant, Adaptive Up, and Adaptive Down. Although there were no differences in plan quality, there were differences in the number and type of factors considered. The Fixed group considered significantly more irrelevant factors and relevant factors, and more factors overall. They did not consider more key factors. Thus is appears that the Fixed group considered an abundance of medium and low-quality information and were not discriminating which information was highly diagnostic. The Fixed progression of difficulty (as opposed to adaptive progression) may inhibit the benefits of diagnostic evaluation in a complex task by not helping participants learn to identify key factors.

In the constant condition, the difficulty was always high and participants had to learn the entire scope of the task at once. Although the scenarios are more involved and may take longer, there are likely to be fewer scenarios in this type of sequencing because participants don't have to "work their way up" to the more difficult scenarios. In a sense, the Constant condition was handicapped with fewer scenarios and learning opportunities. However, this did not negatively impact asset score. The Constant training condition took the least amount of time to train and was significantly faster than either adaptive condition. Compared to the Constant condition, the Fixed condition took 14% longer to train and the Adaptive Up and Adaptive down took 57% and 66% longer respectively. Therefore, the Constant condition was the most *efficient* condition. The training time was significantly shorter in the Constant condition, without sacrificing plan quality.

## CONCLUSIONS

When training complex cognitive skills, having scenarios at a constant high difficulty could benefit training efficiency. It has the potential to reduce training times as well as development times because developers can produce fewer scenarios and do not have to outline and implement a set of complicated adaptive training rules.

Monitoring the decision process is a key component for understanding training pitfalls and also providing more targeted feedback in complex environments. It is not just the *amount* of information that is important for solving

complex tasks; homing in on diagnostic information was associated with better performance. Participants who made better plans considered more key factors and less irrelevant factors. This finding suggests a potential for cost reduction during the training process. A focused approach that targets salient features of complex tasks could serve as an efficient method of eliminating excessive cognitive load while maintaining effective task performance.

More specifically, it is possible to use the information considered to adapt task difficulty. For example, if a trainee is ignoring an entire display or an entire category of information, the task difficulty could be decreased until the trainee learns to consider that information. Thus the task difficulty is not just based on task performance but on decision process. This could reduce situations where trainees "get lucky" and get a correct answer without understanding the answer and using appropriate information. Adapting task difficulty based on decision process is viable option, especially in cognitively complex tasks, and warrants further research.

## ACKNOWLEDGMENTS

## REFERENCES

Barrett, G. V., Greenawalt, J. P., Thornton, C. L., & Williamson, T. R. (1977). Adaptive training and individual differences in perception. *Perceptual and Motor Skills, 44*, 875-880.

Landsberg, C. R., Astwood, R. S., Jr., Van Buskirk, W. L., Townsend, L. N., Steinhauser, N. B., & Mercado, A. D. (2012). Review of adaptive training system techniques. *Military Psychology, 24*, 96-113.

McDermott, P. L., Carolan, T. F., & Gronowski, M. G. (2012). Application of worked examples to unmanned vehicle route planning. Paper No. 12292. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., & Van Gog, T. (2009). Principles for designing effective and efficient training of complex cognitive skills. In Durso, F. T. (Ed.), *Reviews of human factors and ergonomics, 5* (pp 166-194). Human Factors and Ergonomics Society.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*, 1-48.

Salden, R. J., Paas, F., Broers, N. J., & van Merrienboer, J. J. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science 32*, 153-172.

Spain, R. D., Priest, H. A., & Murphy, J. S. (2012). Current trends in adaptive training with military applications: An introduction. *Military Psychology, 24,* 87-95.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science, 12*(2), 257-285.

Van Merriënboer, J.J.G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology 20*, 343-352.

Wickens, C. D., Hutchins, S., Carolan, T.F., & Cumming, J. (2012). Effectiveness of Part Task Training and Increasing Difficulty Training Strategies: A Meta-Analysis Approach. Submitted by Alion Science and Technology to the US Army Research Institute for the Behavioral and Social Sciences for consideration as a Technical Report, Contract #: W91WAW-09-C-0081.Wightman, D. & Lintern, G. (1985). Part task training for tracking and manual control. *Human Factors, 27*, 267-284.

Wexley, K. N., & Latham, G. P. (1991). *Developing and training human resources in organizations.* New York, NY: Harper Collins.