# Investigating the Impact of Training on Performance

**Thomas Carolan, Patricia McDermott, Shaun Hutchins, Chris Wickens**
**Alion Science & Technology**
**Boulder, CO**
**tcarolan@alionscience.com, pmcdermott@alionscience.com, shutchins@alionscience.com, cwickens@alionscience.com**

**James Belanich**
**US Army Research Institute for the Behavioral and Social Sciences**
**Arlington, VA**
**james.belanich@us.army.mil**

## ABSTRACT

How do training methods impact subsequent task performance by a trainee? Which combinations of training methods are more effective for training different types of tasks? Decisions on how to efficiently train a workforce require an understanding of the factors that impact training effectiveness, cost effectiveness, need for periodic retraining, and training transfer. This paper describes a research project that identifies relationships between training methods, task characteristics, and outcome measures that can provide an evidence base to guide training design decisions. The focus is on training methods that promote retention and transfer of the complex cognitive skills involved in technology related military task domains. During the initial phase of the research we conducted a broad review of the published research literature on training effectiveness and training methods. Based on this review an organizing framework was developed to consider training methods within the context of training design decisions. Training method specific meta-analyses were conducted to generate training effectiveness estimates based on analysis of primary studies. Given the emphasis on complex cognitive tasks and training transfer, six methods associated with managing trainee workload and/or promoting transfer were selected for the initial phase of the meta-analysis research. We discuss meta-analysis results for two training methods, worked examples and exploratory learning.

## ABOUT THE AUTHORS

**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.

**Dr. James Belanich** is a research psychologist and the Chief of the Program, Budget, and Strategy Office for the U.S. Army Research Institute for the Behavioral and Social Sciences. His research has focused on improving the effectiveness of training methods and tools. He has authored numerous research reports, journal articles, and book chapters on training methods, distributed learning, game-based training, instructional technology, and adaptive technology. He received his Ph.D. in Psychology – Learning Processes from the City University of New York.

**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 the design of systems and training to support decision making and situation awareness.

**Shaun Hutchins** is a Senior Human Factors Engineer with Alion Science and Technology in Boulder, CO. He studied Engineering Psychology at New Mexico State University where he earned his MA in Experimental Psychology and Experimental Statistics. Shaun is pursuing his PhD in Research Methodology at Colorado State University.

**Chris Wickens** has been a senior scientist at Alion Science & Technology since 2005. At the University of Illinois he was on the faculty of Psychology from 1983 – 2005. In 1984 he was promoted to full professor and jointly appointed at the Institute of Aviation, where he served as professor of aviation human factors. He received his PhD in experimental psychology from University of Michigan in 1974 after serving 3 years in the US Navy. His interests are in attention, multi-tasking, display design and computational models of cognition.

# Investigating the Impact of Training on Performance

**Thomas Carolan**
**Patricia McDermott, Shaun Hutchins, Chris Wickens**

**Alion Science & Technology**
**Boulder, CO**

**James Belanich**

**US Army Research Institute for the**
**Behavioral and Social Sciences**
**Arlington, VA**

## INTRODUCTION

Decisions on how to efficiently train a workforce require an understanding of the factors that impact training effectiveness, cost effectiveness, need for periodic retraining, and training transfer. The factors that influence training design and learning outcomes include individual, training, and organizational characteristics (e.g. Alvarez, Salas, and Garofano, 2004; Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991; Baldwin and Ford, 1988). This paper discusses an ongoing research project to identify evidence-based guidelines for the effectiveness and efficiency of different training methods with application to Army relevant tasks. The initial focus is on training needs for technology related task domains and training methods that support the acquisition, retention and transfer of cognitive skills involved in complex task domains. In particular the research is aimed at identifying quantitative relations between training methods, task characteristics, and outcome criteria that can provide evidence to inform training design decisions.

The overall research approach includes a literature review and meta-analysis phase, an experimental phase, and development of a database to organize the research findings. The database can potentially inform decision aids for choosing training methods/tools that would be effective for a particular set of circumstances. The literature review involved a systematic review/synthesis of the training research literature to organize and analyze empirical findings (e.g., Cooper & Hedges, 2009). The research synthesis goals included (1) identifying evidence to support generalization for training development guidelines and (2) identifying areas where the research evidence is limited or ambiguous and additional research is needed.

The literature review provided evidence that the relative effectiveness of many training methods are moderated to a greater or lesser degree by interactions among outcome criteria, task factors, and individual differences in ability or experience. The meta-analysis was designed to systematically evaluate the effectiveness of training methods in the context of these key moderating factors. The meta-analyses results, in addition to providing empirical results on the impact of training methods on subsequent performance of trained skills, also brought into focus training method specific research questions addressing gaps in empirical evidence on which training methods were more/less effective in particular training situations. The results from the review and meta-analysis determined subsequent experiments that were designed to address training method specific research questions as applied to basic and complex cognitive skills.

## CONCEPTUAL FRAMEWORK

The conceptual framework for the research was driven in part by the evidence database objective. Given a training need, what are the factors to be considered in selecting a combination of training delivery methods and media to maximize the effectiveness and efficiency of training? Three categories of decision factors are considered: What needs to be trained (task factors), who needs to be trained (individual factors), and the outcome objectives. In addition to these training objectives, situational factors, such as cost and availability, may limit the training approach options available for consideration. Training approach selection then considers the relative effectiveness of training methods and media given training objectives and context. Training researchers emphasize the importance of a needs analysis process in determining training objectives which should then drive training design decisions (e.g., Tannenbaum et al 1991; Arthur et al 2003; Alvarez, et al 2004). Using a needs analysis to inform training design decisions implies an organizational framework that supports specifying training objectives and constraints and assessing evidence indicating the conditions under which particular training methods are more/less effective.

The qualitative literature review focused initially on training effectiveness models, meta-analyses, and major reviews to identify the extent to which interactions exist between outcome criteria, task factors, and individual differences in moderating the effectiveness of training outcomes and training methods.

**Outcome Criteria**

The desired outcomes of skills training typically involve acquisition, retention and transfer of skills to performance in an operational environment, while minimizing training time and cost. To what extent should tradeoffs in emphasis among these objectives influence training design decisions? Measures of the effectiveness of a training approach can vary based on evaluation criteria (e.g., Alliger et al, 1997; Arthur et al 2003; Kraiger et al., 1993). Two issues are particularly relevant to training design and transfer effectiveness. The first is that similar training methods can produce different results when measured using learning, retention, or transfer outcome criteria. Some methods, typically highly structured training methods (e.g., low variability practice, complete guidance, immediate corrective feedback), that are effective for skill acquisition as measured by learning criteria are less effective or even detrimental to transfer performance (Schmidt & Bjork, 1992). In addition, there is evidence that some less structured training methods that require more learner effort can yield higher performance during transfer, especially transfer to new problem situations (e.g., Schmidt & Bjork, 1992; Van Merrienboer, Kester & Paas, 2006; Bell and Kozlowski, 2008). This has been referred to as the transfer crossover effect (Bell and Kozlowski, 2008) or transfer paradox (van Merrienboer, Kester and Paas, 2006).

A second issue is that differences in the definition and measurement of transfer may have implications for evaluating training effectiveness. Transfer is defined in terms of the extent to which knowledge and skills learned in one context influences benefits learning or performance in another context, with the extent of transfer based, primarily, on the similarity of training and transfer tasks and performance environments (Wickens & Hollands, 2000). Recent research has characterized the relative similarity between learning and transfer tasks as a near-far transfer distinction (e.g., Barnett & Ceci, 2002). Near transfer involves application of skills to a task or situation very similar to the learning tasks (analogical tasks). The transfer task can be 'far' along a number of dimensions (Barnett & Ceci, 2002). It can involve transfer to a different performance environment or to a more complex task or higher level skill (vertical transfer); such as in part task training. Far transfer can also involve adapting the learned skills to structurally different situations or problems. Emerging Army task domains and mission environments require flexibility to adapt learned skills to new tasks, missions, situations, equipment and environments (Auffrey, Mirabella, & Siebold, 2001). The technology may be changing so quickly that the training community does not have access to the latest equipment or equipment versions. Trainees learn on available equipment and transfer what they learn to the actual systems in the field. For technology-related task domains, the ability to transfer learning from one version of a system to another can be an important training effectiveness factor. Transfer of training is therefore a more broadly defined concept that requires more specific transfer objectives to inform training design. Although this area is a rich research topic, there has not been a systematic review of the impact of various training methods on transfer performance in general, or comparison of methods for more specific near and far transfer objectives. The meta-analysis phase of this research was designed to address the effectiveness of training methods based on different transfer criteria.

**Task Characteristics**

Of the three factors broadly considered under training objectives, the role of task or skill type has played the least prominent role in training effectiveness and transfer models. Task characteristics include factors such as task type, task complexity or difficulty and the types of knowledge and skills required. Task type implies a typology that characterizes target tasks in terms of component skills that potentially benefit from different training methods. For complex cognitive tasks the importance of cognitive task analysis as part of the training needs analysis has been advocated by numerous researchers (e.g., Frederiksen and White, 1989; Goettl & Shute, 1996), suggesting that different cognitive skills respond better to different training methods (e.g., Seamster et al. 1997). However, our review of the training effectiveness research found few studies that explicitly compared the effectiveness of training methods for different task or skill types. There is some evidence that psychomotor, cognitive and interpersonal skills moderate the effectiveness of different training delivery methods (Arthur, Bennett et al., 2003). Research on training specificity (Healy, Wohldmann et al., 2008) suggests limited transfer for skill based tasks and greater transfer for declarative knowledge based tasks. Differences in transfer effectiveness of training methods have also been reported for open skills, such as interpersonal skills, and closed skills that are highly structured and proceduralized (Blume et al., 2009). Other than these very general findings, there has been limited research comparing training methods for different task or skill types. We began to address this general research question in the meta-analysis phase through a detailed moderator analysis of task and skill types used in the studies analyzed for each method.

**Task Complexity/Difficulty**

A growing body of research suggests that task complexity/difficulty is a factor in the relative effectiveness of methods for training complex cognitive skills. Task characteristics such as the interactivity of task elements, the degree of task structure, the potential number of solution paths and possible solutions (e.g., Paas & van Gog, 2009) contribute to task complexity. From the perspective of cognitive load theory (e.g., Sweller, 1988; Van Merrienboer, Kester & Paas, 2006; Paas & van Gog, 2009) these intrinsic task characteristics require more cognitive effort on the part of the learner. Two issues related to task complexity/difficulty and relevant to training design decisions are (1) the organization of the component skills and their interactions have implications for training methods (e.g., Goettle and Shute, 1996) and (2) training methods that are effective for promoting retention and transfer for simple tasks are not always effective for complex tasks. There is a relatively sizeable and growing body of research that investigates methods for training complex cognitive skills. However, we have not found a meta-analysis that addresses the training of complex tasks or that uses task complexity as a moderating variable.

**Individual Factors**

Emerging military task requirements, such as new digital technology, often result in the need for training approaches that are effective and efficient for a diverse military population with a range of ability and experience levels. Individual differences in learner characteristics such as aptitude and experience often moderate the effectiveness of various training methods and are important factors in selecting a training approach. The "expertise effect" is related to differences in the effectiveness of task complexity reducing strategies when used for low and high experience individuals. Complexity reducing strategies can negatively impact performance for high experienced trainees (Paas & van Gog, 2009; van Merrienboer Kester & Paas, 2006). There is also a body of research on aptitude by treatment interactions that indicate the effectiveness of training interventions differs depending on learner aptitudes. For example, lower levels of ability tend to benefit from structured lessons and higher levels of ability benefit from less structured training (Bell and Kozlowski, 2008). There may be an impact of skill type on aptitude effects. Arthur et al. (2003) did not find evidence to support aptitude by treatment interactions. Blume et al. (2009) found that skill type modified the relationship between cognitive ability and transfer. They found a small, negative relationship between cognitive ability and the transfer of open skills and a moderate positive relationship between cognitive ability and the transfer of closed skills.

**Training Approach**

Our literature review did not reveal a common taxonomy of training methods, and similar methods are often used to achieve different training goals and/or operationalized differently depending on the research objectives (Clark, 2009). Training methods can be considered in the context of high-level training design dimensions. Given the task, trainee population, constraints and outcome objectives, how important are each of these dimensions and for each category what are the more effective implementation methods? Training method categories include: methods for structuring and guiding training, methods for managing training task complexity, methods for adaptive training, methods for promoting active learning, and methods for learner interactivity. In addition, training methods can be organized around a basic training design process consisting of a set of common design components. Training design decisions address each of these training components and particular training methods are operationalized in terms of one or a combination of elements.

Methods for managing training task complexity often involve training task selection and sequencing methods. *Training task selection* involves selecting methods for decomposing or simplifying the target task into one or more training sub-tasks. In this category we consider methods that manage training complexity by modifying the target task itself. These include methods such as: Part task training, whole task methods that systematically modify the task or task focus for training, some training wheels strategies and simulation/simulator fidelity. *Task sequencing methods* can be fixed, adaptive, learner controlled, advisory, or shared. Fixed can include different part–whole sequencing strategies.

Methods for structuring and guiding training are typically implemented as instruction, guidance during practice, feedback, or some combination. *Instruction* component involves methods for presenting the task content to the trainee and can include overview material, task concepts (declarative knowledge), task procedures, strategies, demonstration, modeling, guidance for exploration and discovery, etc. *Practice* methods and variables include: spacing, variability, context and amount of practice; and can be problem-based, include group interaction, or mental. *Guidance during practice* includes prompts, hints and scaffolding approaches that provide information to support task performance. Guidance can include content-based prompts, metacognitive coaching, directive guidance, conceptual hints etc. Guidance can be solicited only or unsolicited. Also included are strategies for reducing or increasing guidance. *Feedback* involves information on performance provided after the performance. It is useful to distinguish between guidance and feedback. Feedback methods can include type of feedback: knowledge of results or product, process errors, correctional feedback (correct outcome or process), and can be immediate (after each step) or more delayed. These design elements are not meant to be exhaustive and can involve overlapping methods; however they provide a reasonable framework for organizing training methods in the context of training design decisions and a problem space to guide meta-analysis and experimental research.

## META-ANALYSIS

Training method specific meta-analyses were conducted to generate quantitative estimates of the relative effectiveness of targeted training methods and the factors that moderate effectiveness. Given the emphasis on complex cognitive tasks and training transfer, and current issues in the research literature, six methods associated with managing training difficulty and/or promoting transfer were selected for the initial meta-analysis research. The methods evaluated include: scaffolding, part task training, error prevention or training wheels, adaptive sequencing / increasing difficulty, learner control, and exploration-based learning. The first four training methods combine strategies for managing the difficulty of the target task during training and sequencing from lower task difficulty to the full complexity. Learner control and exploratory learning are considered active learning methods that have been linked to increased transfer effectiveness when combined with other training strategies. Each of these training methods has had mixed results with respect to training effectiveness and their effectiveness is thought to be moderated by other variables. The meta-analyses address an important gap in the training effectiveness and learning transfer evidence base.

The meta-analysis search was broad and inclusive. Minimum criteria included normal adult populations, a training strategy manipulation versus a control condition, and a measure of the transfer of training. Transfer tests were sub-defined to include near transfer or far transfer tests of knowledge and performance. We sought to both confirm and explore moderating influences on effective training interventions. A comprehensive coding scheme was developed to capture relevant qualitative method and moderator variables in addition to the quantitative 'effects' data. Potential moderator variables included: task type, task complexity, type of skill, trainee experience, trainee ability, type of outcome measure, transfer type, and strategy-specific variations in training methods.

Two complementary measures of treatment effect were used, the conventional effect size metrics (*d'*, *g*) (Rosenthal, 1991) and a transfer ratio score (TR) based on the ratio of transfer performance achieved by the chosen treatment, to the performance achieved in the control condition. In this paper, we present the main and moderator effect results using a conventional effect size metric (Hedges' g). A more detailed discussion of the meta-analysis including methodological differences and rationale for computing complementary measures can be found in Wickens, Hutchins, Carolan and Cumming (2011). In this section, we provide brief overviews of two training strategies, exploratory learning and worked examples, summarize the meta-analysis results for each approach, and discuss research implications.

### Exploration-based Learning

There is disagreement and conflicting findings on the advantages and disadvantages of highly structured and directed versus less structured, more learner controlled training methods for learning, retention and transfer outcomes (e.g., Kirschner, Sweller & Clark, 2006). Highly structured training methods typically involve design elements that include direct instruction, fixed low to high difficulty sequencing, massed and low variability practice, continuous guidance and support, and immediate corrective feedback. Clark (2009) reviews evidence that directed guidance methods providing complete task descriptions of when and how to perform procedures is more effective and efficient than more exploratory/discovery methods, is applicable to complex cognitive tasks, and supports transfer, including adaptive transfer. While the research evidence supports highly structured methods as effective for training to very task specific learning objectives, there is evidence of a transfer crossover effect indicating that highly structured training methods often negatively impact transfer of learning to new problem situations whereas less structured training methods that produce lower performance during training often yield higher performance during transfer (e.g., Schmidt and Bjork, 1992; van Merrienboer, et al, 2006; Bell and Kozlowski, 2008). Emerging evidence suggests that more exploration-based learning approaches can improve learning outcomes and facilitate adaptive transfer (e.g., Bell and Kozlowski, 2008). The general hypothesis is that exploratory learning promotes metacognitive activity and self-regulation of learning skills (Bransford et al., 1999) which are thought to be critical for complex skill development and adaptive transfer.

Exploratory learning methods involve active learner exploration in the task environment with the level and type of guidance as the primary modifying variable. The terms exploratory and discovery learning are often used to refer to training that provides little or no external structure or guidance to direct active exploration in the task environment, hence unguided or learner guided exploration. Between the unstructured and highly structured approaches are exploration based approaches with various levels and types of guidance. While the research evidence suggests that fully unguided exploratory learning is less effective than training with some degree of structure and guidance, there is evidence that minimally guided strategies can be more effective for transfer. In a recent meta-analysis, Keith & Frese (2008) investigated the effectiveness of error management training (EMT) on learning and transfer for skill-based training, in particular software operation skills. The meta-analysis included studies comparing EMT with alternative training methods including purely exploratory training and procedure based training. They report a significant positive mean effect for EMT with moderator analyses indicating larger effect sizes for post training transfer than for within-training performance, and for tasks that were structurally different from training tasks and require adaptive transfer than for near transfer tasks that were similar to training. Between minimally guided and direct guidance approaches there are a range of strategies for providing external direction to engage trainees in systematic, preplanned exploration (Bell and Kozlowski, 2008). Examples include exploration following a lesson or behavior modeling (e.g., Frese et al., 1991) and exploration guided by particular problem solving goals (e.g., Dyer, Singh & Clark, 2005). Goodwin (2006) cites a number of Army studies indicating that guided exploration has been more effective than other conditions tested for learning outcomes. Our literature search did not uncover meta-analysis research systematically investigating the effectiveness of exploratory learning with different levels and types of guidance.

**Exploratory Learning Meta-analysis**

Given our research objectives and conceptual framework, the goal of the exploratory learning meta-analysis is to generate evidence that can inform training design decisions with respect to the relative effectiveness of more learner guided exploration approaches versus more externally guided or directed approaches given a particular training situation.

The primary research question addresses the benefit or cost of more structured and guided training methods relative to less guided more learner exploration and discovery based approaches. Based on the research literature the expectation is that outcome criteria, task factors, and individual differences will significantly moderate the relative effectiveness of more externally directed and more learner exploration based approaches. A second objective is to estimate the impact on transfer performance of different amounts and types of training guidance, relative to more learner guided exploration based learning. As discussed in the previous section, from a training design element framework, methods for structuring and guiding training can be implemented as pre-practice instruction, guidance during practice, feedback, or some combination. Coding of the guidance details for each study is still in progress. In this paper we present partial results focused on the relative effectiveness of more versus less guidance, based on coding the level of guidance provided in each study treatment and control condition.

Our meta-analysis literature search and selection process yielded 31 exploratory learning research studies that provided statistical information necessary for the effect size analysis. The 31 papers provide a total of 135 data points for the analysis, including multiple sub groups, treatment comparisons and outcome measures. Training conditions were coded as directed guidance, guided exploration, minimally guided exploration, or full exploration. Since each study is a contrast between two levels of learning guidance, coding was relative in some cases to assure different levels for each condition. Conditions were coded as directed if explicit step by step instructions were provided, a rule to be leaned was provided with examples, or direct guidance, prompts, or immediate corrective feedback was provided during practice. Conditions were coded as guided if instruction included demonstration, worked example, conceptual framework, or solution strategies, and/or if unsolicited procedural, conceptual or metacognitive prompts or corrective feedback were provided during practice. Conditions were coded as minimally guided if prior instruction was limited to providing high level problem structure or task overview, practice includes trial specific learning objectives or tasks to perform, or limited conceptual or coaching guidance was available if solicited. Conditions were considered learner guided or full exploration if prior instruction was orientation and overview only, there was no learning content related instructor interaction, and access to a manual was allowed but discouraged. For each comparison, the more guided condition was coded as ‘treatment’ and the more exploratory as ‘control’.

The overall mean effect size across all comparisons and outcomes measures was positive (Hedges g = .15) indicating a small but significant (p>.001) overall benefit for more guided practice or exploration relative to less structured and guided exploration approaches. Tests for heterogeneity indicate significant variation with approximately 83% of the unexplained variation due to heterogeneity. These results imply either mixed effectiveness of exploration-based learning, or significant moderators of the effect of exploration-based learning. In the following discussion all effect sizes presented are significant unless stated otherwise and all effect sizes are Hedges g.

**Outcome measures.** Of the 135 outcome measure data points, 27 were declarative knowledge based measures and 108 were performance based tests (including paper and pencil tests that required problem solving). The average effect size across the knowledge based measures was positive (.54) indicating a large benefit for more guided practice or exploration over less guidance. For the performance based measures, the effect was also positive (.093) indicating a small average benefit over all transfer measures for more guided practice or exploration over less guided approaches.

The outcome measures were also categorized in terms of the type of transfer. Three categories were coded: the transfer task was either (1) identical to the training task, (2) similar to the training task (Near Transfer), or (3) required adapting learned skills to different and/or more complex tasks or scenarios (Far Transfer). In addition the measure could be a second test of the transfer task to measure retention. The effect sizes for the declarative knowledge-based measures were large and positive for training identical (.928) and near transfer (.441) tasks. There were no declarative knowledge far transfer measures. For the performance measures, the breakdown of effect size by transfer type follows the expected transfer crossover effect with training identical transfer (.309) and near transfer (.105) benefiting from more guidance, and far transfer (-.136) benefiting from less guided exploration.

The performance measure data includes multiple performance measurement types from some studies; a breakdown by outcome measure type provides a more accurate comparison across studies and transfer types. Fifty-six accuracy or error based performance measures yielded a close to zero average effect size but a clear crossover over effect across transfer task type. Training identical transfer benefited from more guidance (.293) but near transfer (-.104) and far transfer (-.345) effect direction and size suggest more benefit for transfer performance from less guided exploration based training. Time based measures (time to completion, time or trials to criterion) represented 12 data points (7 near transfer and 5 far transfer). The relatively large average effect size for near (-.32) and far (-.61) transfer suggest that for these studies performance speed on transfer tasks benefited from less direct guidance during training. The analysis of efficiency measures (e.g., number of procedures to complete task) indicated a benefit for more guidance (.638) for the 9 near transfer measures and a benefit for less guidance (-.454) for the 1 far transfer case. In summary, the analysis of outcome measure type moderator effects from this meta-analysis supports the transfer crossover effect for more and less guided training strategies when the transfer measures are performance accuracy or transfer task learning or performance speed, but not for declarative knowledge measures or performance efficiency transfer measures.

**Task and skill types.** Studies were coded for task domain type and skill type. The task domain coding is meant to categorize similar tasks at a higher level than component task type to support mapping to military tasks. The most prominent task domain in the research set is software productivity applications (electronic search, email, presentation, word processing, etc.,10 studies), next is science laboratory tasks (5 studies), and "lab problems" (3 studies), defined as problems developed for research (McDaniel & Schlager, 1990) and here include logic, transformation and puzzle type problems that typically involve learning rules from examples. All other task domains have 1 or 2 studies. To assess the potential for differences in the effectiveness of more exploratory training for different types of tasks, it is useful to compare the effect size and direction on combined near and far transfer measures across task domains, since far transfer alone provides a more limited set of data points across studies. The average transfer performance over near and far transfer was significant and negative, indicating an overall benefit for less guided exploration, for software applications (-.176) and lab problems (-.695). The science lab task category average effect size data indicates a small but not significant cost to transfer performance due less guidance during training. While there are very likely other factors contributing, the task moderator results suggest that the effectiveness of exploration based training approaches may also depend on task related factors. The skill type moderator analysis suggests a possible explanation. For tasks that involve primarily procedure learning skills, the average transfer performance over near and far transfer was significant and negative (-.158), indicating an overall benefit for less guided exploration. For tasks that involve more problem solving, the average transfer performance over near and far transfer was significant and positive (.172), indicating an overall benefit for more guided or directed training.

**Individual differences.** Considering only the studies where the sample was college students or graduates, the average overall effect size for guidance on transfer was near zero. There were only two studies that explicitly compared college and non-college sub groups. The overall effect size (.961) for the combined non-college sample indicates a substantial benefit to transfer (combined near and far) performance from more guided training methods relative to less guidance. There were only a few studies that explicitly compared high and low ability groups and the average effect size for guidance methods on transfer was not significant for either ability level sub group. While the non-college data is based on only two studies, the large difference between the non-college and college group measures suggests that more research is needed that compares the effectiveness of training approaches on non-college population samples.

**Levels of guidance.** As a first step toward identifying the relative differences in transfer effectiveness for different levels and types of guidance, we looked at the effect of different levels of guidance on training and transfer performance. The four levels of guidance provided six levels of guidance comparisons: directed vs. guided, minimally guided, and full exploration; guided vs. minimally guided and full exploration, and minimally guided vs. full exploration. For performance measured on the training identical transfer task, where the overall effect favored more guidance, directed training was significantly more effective when compared to exploratory-based training regardless of the level of guidance for exploration, and more guided exploration tended to be more effective when measured by training task performance than less guided exploration.

For performance measured on near transfer, where the overall effect also favored more guidance, comparison of exploration-based conditions indicated a tendency to favor the more guided exploration over the less guided approaches but the effect sizes were not significant. Directed training was more effective than guided exploration, except for the directed versus full exploratory comparison, where exploration was more effective. For far transfer, where the overall effect favored more exploration, transfer performance favored the more exploratory condition at all levels except the comparison between directed and guided exploration which significantly favored the directed approach.

When the exploration and the training were integrated, that is there was not a separate instructional session prior to exploration for either the more guided or less guided condition, the overall effect favored the more exploratory training condition (-.156). Transfer performance benefited more from the more guided training, when the more guided condition consisted of instruction followed by exploration (.375) or when both conditions consisted of an instruction session followed by an exploration or practice session (.243).

The exploratory learning meta-analysis results summarized here suggest that the most effective level and type of structure and guidance for learning may depend on task and individual factors and the specific learning transfer objectives. The overall findings suggest that on average more structure and guidance during training benefits training task performance and knowledge transfer and more exploration based training benefits far transfer task performance. The task and skill type moderator analyses suggests a potential moderating effect based on task factors, with more procedural tasks, benefiting from more exploration-based approaches and tasks that involve more problem solving, benefiting from more guided or directed training. The skill transfer benefits of less structured, more active learning under conditions where the trainees' available cognitive load is not exceeded is still an open research question.

## Worked Examples

The 'worked examples' strategy is considered a type of error prevention strategy designed to provide the correct steps to take in a problem-solving process. Typically, this strategy involves the presentation and/or demonstration of a fully or partially worked problem with 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 et al., 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 et al., 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 our meta-analysis, the worked examples method was analyzed as a type of error prevention strategy. The literature search and selection phase yielded 27 studies with a total of 71 data points for the effect size analysis. The main research question addressed the benefit of using worked examples as an error prevention strategy. Using Hedges' g, and a fixed effect model, the effect size (.208) indicates a significant transfer benefit for using worked examples. Of these 71 data points, 50 involved only a worked example technique, the others used worked examples in conjunction with other error prevention techniques such as prompting. Worked example only conditions yields a larger effect size (.307) than worked examples with prompting (.035). We categorized the worked example tasks according to the primary types of problem solving skill required: cognitive procedures, declarative knowledge, quantitative skills, general problem solving (reasoning) skills and spatial reasoning skills. For tasks where the problems required combining two types of reasoning skills (typically general reasoning plus quantitative or spatial) we coded as requiring both. One potentially important finding is that worked examples benefited problems that primarily involved a single skill type, for example general reasoning (.47) or quantitative reasoning (.44); but did not benefit problems requiring a combination of two skills– general reasoning & quantitative reasoning (-.03), general reasoning & spatial reasoning (-.31). One possible explanation is that the number of skills required is an indicator of the complexity of the problem type. Other factors also suggest that using worked examples to guide learning may not be effective for complex problems. For the 'type of transfer' moderator, worked examples were effective when the transfer problem was identical (.23) or similar (.48) to the training problems, and when the transfer problem required a new problem type (.33) but not when transfer was to a more complex problem (-.07). There were few studies where ability or experience was addressed as moderating factors. For the single study that manipulated ability the effect of WE was stronger for low ability (.18) than for high ability (.02) trainees. Worked examples benefited both low (.41) and high (.50) experience levels. 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 (.83) than when used during the practice phase (.10) or during both phases (.15). This suggests worked examples may be more effective for providing a problem schema for use as initial guidance. 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 results also point to a number of areas for further research relevant to military training needs. One area is application to a broader set of problem types. The meta-analysis suggests that worked examples are effective methods for structuring problem solving to benefit performance on near transfer tasks but not for far transfer tasks. Research combining worked examples within a more active learning framework would provide evidence for the potential benefit of worked example strategies to more adaptive transfer. In an experiment currently in progress, we extend the worked example method to unmanned vehicle route planning problems and investigate their use within a more directed and a more exploratory guidance approach. Transfer tasks are used to evaluate the effectiveness of the worked example methods in supporting transfer from structured training tasks to less structured planning problems requiring more inferential reasoning and to performance in a new task environment.

## EXPERIMENTAL RESEARCH

The literature review and meta-analysis brings into focus a number of areas where the research evidence is limited, results are mixed, and/or there are moderating factors still to be investigated. The results of the meta-analyses are being used to define research questions within the context of Army relevant tasks. The conceptual framework was applied to consider training design research questions within the context of the task, individual and outcome objective moderating factors. The limited research findings comparing the effectiveness of training methods for different task or skill types motivated a task/skill type driven experimental approach, in the sense that the objective is to address training method research questions with application to particular Army relevant task/skill types and within the context of a common Army task domain. Experimental studies were initiated to augment evidence generated from the literature review and meta-analysis through an investigation of the effectiveness of combinations of training methods for learning particular skill sets in the context of Army task domains. The experimental focus is on simulation-based training and blended delivery approaches that combine computer simulation environments with classroom instruction, with the guiding research question: If you are in a classroom and have a simulation environment available, what is the best way to use it to train difficult tasks?

Two common design elements provide a common structure to investigate research questions. First is a set of outcome criteria that provides a systematic approach for investigating tradeoffs between training methods that are targeted at very specific task objectives and those that are targeted at flexibility for transfer of knowledge and skills. Dependent measures for each study are designed to address effectiveness of the training interventions across the range of outcome criteria: skill acquisition, training efficiency, skill retention, and transfer of skills. Initial acquisition and near transfer measures provide a comparison baseline for training effectiveness and training efficiency trade-offs. Near transfer tasks are used to assess the maintenance of trained skill after a retention interval. Far transfer tasks are used to assess transfer to a new problem with structural differences, greater complexity, or in a new performance environment. Each experiment includes three types of transfer tasks: a near and a far transfer test after the training session and a delayed transfer task that involves transfer to a new performance environment. The second common design element addresses the role of individual differences in ability and/or experiences in moderating the effectiveness of any training interventions. Participants are sampled from both college students and from vocational school or non-college sources to provide a comparison

Three experiments were designed to compare training methods for the types of basic and complex cognitive skills required to operate and employ robotic systems. Employing robotic systems involves a range of skills from basic procedural skills needed to effectively operate complex digital systems, to more complex cognitive skills required for planning, teleoperating, navigating and monitoring unmanned systems. Each of the three experiments focused on a different set of skills. Experiment 1 addressed basic cognitive procedural skills needed to perform tasks using a digital interface to interact with unmanned vehicles. Experiment 2 focused on higher order cognitive planning skills. Experiment 3 combined visual-spatial reasoning, perceptual-motor coordination and timesharing involved in teleoperation tasks.

One practical training question, in the context of a phased approach to training, involves the potential training efficiency benefit to having trainees learn and practice the basic digital system operating procedures on their own, before starting a training program on the more complex aspects of employing a digital system interface such as robotic control systems. A research program that uses learner-guided exploration training as the baseline control for evaluating the effectiveness and efficiency tradeoffs of more resource intensive guided or directed training approaches would provide systematic evidence to support training design decisions, given task and trainee factors and outcome objectives. Relatively low complexity cognitive tasks such as HCI procedures, where self-training seems a reasonable approach, provides a test of this strategy.

The first two experiments were designed to investigate the effectiveness of different levels and types of guidance for procedural and planning tasks respectively. Based on the exploratory learning meta-analysis, a more guided training approach may be more effective in promoting training transfer for problem solving tasks such as route planning, whereas a less guided approach may be more beneficial to training transfer for more procedural tasks. In Experiment 1, a training structure variable compared highly structured and less structured exploration-based approaches implemented through combinations of instruction, guidance and feedback design elements. The meta-analysis indicates that worked examples are effective methods for structuring problem solving for near transfer tasks but not for the far transfer tasks. In Experiment 2 the training method manipulation compares the use of worked examples for route planning problems within a more directed and a more exploratory guidance approach.

## SUMMARY

This paper provides an overview of an ongoing research project to identify evidence-based guidelines for the relative effectiveness of training methods to support retention and transfer of cognitive skills involved in complex task domains. The meta-analysis and empirical research framework was designed to understand these moderating factors in order to draw useful guidelines. The meta-analyses of worked examples and exploratory learning research provides some quantitative evidence for conditions under which these training methods may be expected to benefit transfer effectiveness. The results of the meta-analyses are being used to identify research questions relevant to military tasks and strategies to support transfer from training environment and conditions to performance with actual equipment in the field.

## ACKNOWLEDGEMENTS

This work is supported by the US Army Research Institute under Contract #: W91WAW-09-C-0081 to Alion Science and Technology titled Understanding the Impact of Training on Performance.

## REFERENCES

Alliger, G. M., S. I. Tannenbaum, Bennett, W. Jr. Traver, H., & Shotland, A. (1997). A Meta-Analysis of the Relations Among Training Criteria. *Personnel Psychology 50*(2), 341-358.

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*(4), 385-416.

Arthur, W., Jr., Bennett, W., Jr., Edens, P.S., & Bell, S.T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234-245.

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.

Auffrey, A.L., Mirabella, A., & Siebold, G.L. (2001). *Transfer of training revisited* (Research Note 2001-10). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Baldwin, T.T., & Ford, J.K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41, 63-105.

Barnett, S. M. & Ceci, S. J. (2002). When and Where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128(4)*, 612-637.

Bell, Bradford S.; Kozlowski, Steve W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*(2), 296-316

Blume, B., Ford, J., Baldwin, T., & Huang, J. (2009). Transfer of Training: A Meta Analytic Review. *Journal of Management, 36*(4), 1065-1105.

Bransford, J.D., Brown, A.L., and Cocking, R.R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press.

Clark, R. E. (2009). How much and what type of guidance is optimal for learning from instruction? In S. Tobias & T. Duffy (Eds.), *Constructivist theory applied to Instruction: success or failure?* (pp. 158-183). New York, NY: Taylor and Francis.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis (2nd ed.)*. New York: Russell Sage Foundation.

Dyer, J. L., Singh, H., & Clark, T. L. (2005). Computer-based training approaches for training interactive digital map displays (ARI Research Report 1842). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Frederiksen, J., White, B. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 71*, 89-146.

Frese, M., Bodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction, 6*, 77-93.

Goettl, B. P., & Shute, V. J. (1996). Analysis of part-task training using the backward-transfer technique. *Journal of Experimental Psychology: Applied, 2*(3), 227-249.

Goodwin, G. A. (2006). The training, retention, and assessment of digital skills: A review and integration of the literature (ARI Research Report 1864). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Healy, A., Wohldmann, E., Kole, J., Schneider, V., Shea, K., & Bourne, L., Jr. (2008*). Training for efficient, durable, and flexible performance in the military* (Technical Report 1220). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Keith, N. and M. Frese (2008). Effectiveness of Error Management Training: A Meta-Analysis. *Journal of Applied Psychology*

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41, 75-86.

Kraiger, K., Ford, J., and Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78(2,* 311-328.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. Cognition and Instruction, 7, 129-159.

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.

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.

Rosenthal, R. (1991). Meta-Analytic Procedures for Social Research. Rev Ed. Sage Publications, Inc., Beverly Hills, CA.

Sanders, W. R. (1999). Digital procedural skill retention for selected M1A2 tank intervehicular information systems (IVIS) (Research Report 1432). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schmidt, R. A., Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Seamster, T., Redding, R., & Kaempf, G. (1997). *Applied Cognitive Task Analysis in aviation*. Aldershot: Avebury.

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

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology, 76,* 759–769.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review 22*(2), 155-174.

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. and Hollands, J. (2000). *Engineering Psychology and Human Performance*, 3rd Edition, Upper Saddle River, NJ: Prentice Hall.

Wickens, C.D., Hutchins, S., Carolan, T., & Cumming, J. (2011). Investigating the Impact of Training on Transfer: A Meta-Analytic Approach. Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting. 19-23 September, Las Vegas, NV.