Training in process control: Supporting performance in consideration of operator characteristics

Dina Serena Burkolter

Dissertation zur Erlangung der Doktorwürde an der
Philosophischen Fakultät der Universität Freiburg (Schweiz)

Genehmigt von der Philosophischen Fakultät auf Antrag der Professoren
Jürgen Sauer (1. Gutachter) und Dietrich Manzey (2. Gutachter).
Freiburg, den 09.06.2010 (Datum der Thesenverteidigung).
Prof. Thomas Austenfeld, Dekan

To Simon

Summary

This doctoral thesis investigated how training can best support process control performance in consideration of individual differences, i.e. operator characteristics. Process control can be found in industries that control large chemical, energy or thermal processes and is highly safety-critical.

The research question was approached through experimental studies conducted with a simulated process control task. Study I investigated the relationship between operator characteristics and process control performance. Study II replicated and extended these findings by analyzing the relation between the same and additional operator characteristics and process control performance. The focus of study III was on the comparison of three training approaches designed to enhance process control performance. Study IV aimed at applying a novel measure of structural knowledge to test its potential as a training outcome in process control.

Taken together, the results show that effects of operator characteristics and training methods on performance differ with respect to the two main tasks of process control–system control and fault finding. Hence, one needs to consider operator characteristics and design training interventions according to each of the subtasks.

Keywords: Training; Process control; Individual Characteristics; Training Method; Training Outcome

Contents

List of publications of the doctoral thesis

The publications for this doctoral thesis are all written as a first author and include four peer-reviewed articles. The publications will be referred to in the thesis as Study I to IV.

Study I Burkolter, D., Kluge, A., Sauer, J. & Ritzmann, S. (2009). Predictive qualities of operator characteristics for process control performance: The influence of personality and cognitive variables. *Ergonomics, 52*, 302–311.

Study II Burkolter, D., Kluge, A. & Brand, M. (2009). Cognitive and personality variables of operators: individual characteristics and process control performance. In D. de Waard, J. Godthelp, F.L. & K.A. Brookhuis (Eds.). *Human factors, security and safety* (pp. 77–88). Maastricht: Shaker Publishing.

Study III Burkolter, D., Kluge, A., Sauer, J. & Ritzmann, S. (2010). Comparative study of three training methods for enhancing process control performance: emphasis shift training, situation awareness training, and drill and practice. *Computers in Human Behavior, 26*, 976–986.

Study IV Burkolter, D., Meyer, B., Kluge, A. & Sauer, J. (in press). Assessment of structural knowledge as a training outcome in process control environments. *Human Factors*.

Chapter 1

Introduction

1.1 Ensuring safety in work environments

Ensuring safety in work environments is essential –not only for personnel, but also for the organization as well as the environment. Safety is perceived as the functioning of an organization without serious failures or harm for the organization and the environment. Safety is continuously developed from a coaction of intra- and extra- organizational factors such as personnel, technology, or rules (Fahlbruch, Schöbel & Domeinski, 2008).

One measure taken to ensure safety is training. A skilled workforce can yield higher productivity, higher motivation and commitment, and increased safety (Salas, Wilson, Priest & Guthrie, 2006b). Training refers to the acquisition of skills that result in improved performance (Patrick, 1992) and is defined as "the systematic approach to affecting individuals' knowledge, skills, and attitudes in order to improve individual, team, and organizational effectiveness" (Aguinis & Kraiger, 2009, p. 452). As Patrick (1992) states, the important point is that "training is more than just learning" (p. 2), because training is aimed at improving performance on a specified task, while learning refers to the change in a person's behavior in response to a particular situation due to repeated experiences in that situation.

Training of personnel has become an important factor in organizations as can be shown by the resources spent on training. The American Society for Training and Development estimated that U.S. organizations spent $134.1 billion on employee training and development in 2008 (ASTD, 2009). A wide range of different training approaches are employed, but on-the-job training and the lecture method remain the most frequently used. More recent training approaches involve computer- and technology-assisted instruction (Goldstein, 1993; Salas et al., 2006b). Furthermore, individual differences of trainees are of concern in terms of the training design. Training programs need to be designed appropriately to the abilities and the personality of trainees (e.g. Salas et al., 2006b).

Despite these endeavors for personnel training in organizations, "the reality is, all too often, that the training an individual experiences is ineffective and fails to transfer back to the workplace" (Fincham, 1999, p. 36). In this context, Goldstein (1993) refers to the "training struggle". Most organizations do not evaluate their training programs in order to determine their utility. If training is evaluated, then it is done so by means of collecting trainee reactions at the end of the course. However, to assess whether skills were acquired and whether they transfer to job performance, follow-up measurements on the job are needed. Besides the lack of thorough training evaluations, there is also a lack of training needs assessments through task, person and organization analyses (Dipboye, 1997; Goldstein, 1993). However, it is essential to assess training needs, both for personnel and the organization, in order to derive training objectives that enable an effective training program to be designed.

Early endeavors to devise training programs designed to improve safety took place around 1900 with studies devoted to safety training in industries such as mining and railroads (Ford, 1997). More recent advancements stem from the aviation industry. In the 1970s, a series of accidents (e.g. 1977 Tenerife runway collision) implicated human error as a causal factor. Recognizing the significance of human error, the aviation industry introduced special training programs designed to reduce error and improve safety. The aviation industry has been instrumental in the development of safety training called crew resource management training. Crew resource management training has been adopted by other industries with high hazards, including anaesthesiology, air traffic control, the navy, nuclear power industry, aviation maintenance, and the offshore oil industry (Flin, O'Connor & Mearns, 2002; Salas, Wilson, Burke & Wightman, 2006a). In the aviation (civil and military) as well as the nuclear power industry, such crew resource management trainings is mandatory (O'Connor & Flin, 2003; Salas et al., 2006a; Strohschneider, 2008).

In summary, training has become a way of life for many organizations (Salas et al., 2006b). Furthermore, special endeavors are undertaken to ensure safety in high hazard work environments by means of special training programs. However, there still remains the need for a thorough needs assessment and a consideration of individual differences as well as sound training evaluations after training delivery. In this doctoral thesis, training and individual differences in the context of complex human–machine interaction are investigated.

1.2 Contribution of this thesis to human factors as well as work and organizational psychology

Research findings from work and organizational psychology as well as human factors are integrated in this thesis. Traditional research topics from work and organizational psychology –training and individual differences– are investigated in the context of human-machine interaction thereby aiming to gain new insights into human factors. At the same time, the goal is to expand the research field of work and organizational psychology by performing research in a domain that is novel to work and organizational psychology. By presenting the history of both disciplines, their objectives and research fields, these research goals are explained in more detail.

Both the history of work and organizational psychology and that of human factors can be traced back to experimental psychology (for example the work of Wilhelm Wundt). Later on, a more applied psychology began to emerge, as there was a need for a more "usable" psychology (Lane, 2007, p. 244). During World War I, airplanes and tanks became increasingly complex and operator errors started to increase. This led to the military's interest in meeting the demands posed on operators through new systems and determining whether humans were capable of meeting these demands. This work resulted, for instance, in the development of new displays and controls. World War II was the starting point for the divergence of work and organizational psychology and human factors. While work and organizational psychologists were involved with the testing, screening, and classification of recruits, human factors experts were concerned with adapting knowledge of human abilities and limitations to the design of military equipment (Lane, 2007; Muchinsky, 1987). After the two World Wars, members of industry also began to hire human factors experts to design jobs and equipment which led to a further expanding of human factors (Lane, 2007).

Today, human factors investigates how humans accomplish work-related tasks in the context of human-machine interaction while work and organizational psychology has a broader view, in that the concern is with behavior in work situations in general (Chmiel, 2000; Muchinsky, 1987). The two disciplines of work and organizational psychology and human factors do have research fields in common, such as training, effects of stress or team performance. However, they are considered to be distinct entities, because human factors has become a multidisciplinary endeavor that has been influenced not only by psychology, but also by engineering, design, physiology, and computer science (Lane, 2007).

The history of the two disciplines has influenced the topics that are investigated. By inspecting text books from work and organizational psychology (Chmiel, 2000; Fincham, 1999; Nerdinger, Blickle & Schaper, 1999; Weinert, 2004), research topics of the discipline are summarized. Traditional research topics in work and organizational psychology include leadership, motivation and job satisfaction, social interaction including group and intergroup behavior, stress, organizational culture and processes, as well as training and individual differences. These topics were traditionally investigated with regard to work environments such as manufacturing, military, office work or sales (Muchinsky, 1987). Traditional research topics in human factors are, due to its history, more related to cognitive and experimental psychology. They include decision making, attention, perception and information processing, workload and stress, display design, automation, safety, accidents and human error. These topics were traditionally investigated regarding work environments such as aviation, transportation, military or nuclear power (Badke-Schaub, Hofinger & Lauche, 2008; Wickens, Gordon & Liu, 1998; Wickens & Hollands, 2000; Salvendy, 2006).

As mentioned above, training and individual differences are focused on in this thesis. Training and individual differences belong to traditional research topics in work and organizational psychology. They have been investigated in a range of different work environments –as summarized above–, but have rarely been examined in high hazard work environments. Furthermore, most of the research performed on the relationship between personality, ability and performance explored only student performance from grade school through college (Goldstein, 1993).

In the field of human factors, individual differences present a novel topic. In the Handbook of Human Factors (Salvendy, 2006), for instance, contributions to individual differences are included. However, these contributions are all related to design but not to training (e.g. "design for people with functional limitations", "design for aging" or "design for children"). Moreover, the study of individual differences is limited to demographic variables. A review on personnel selection –which is related to research on individual differences– is included in the handbook (Hedge & Borman, 2006), although it is kept on a general level and is not specifically related to high hazard work environments and human-machine interaction. This might be due to the fact that research on personality and ability variables and high hazard work environments are hard to come by, except from the field of aviation (e.g. Wickens et al., 1998).

Training research is approached differently by work and organizational psychology and human factors. For work and organizational psychologists, training is "inherently an applied enterprise in which organizations attempt to change individuals in a way that is consistent with the job requirements" (Quiñones & Ehrenstein, 1997, p. xi), and technological, social, and individual factors are primarily considered. For human factors psychologists or cognitive psychologists, training involves applying principles of learning and skill acquisition and a focus on individual ability, information processing and task characteristics (Quiñones & Ehrenstein, 1997). This doctoral thesis aims at integrating both approaches by drawing on training theories from work and organizational psychology and theories of skill acquisition from human factors research.

In conclusion, this doctoral thesis aims to contribute to filling research gaps in work and organizational psychology as well as human factors. Training and individual differences are studied in a high hazard work environment –process control– integrating research approaches and findings from both disciplines.

The next chapter goes into more detail about training in process control and introduces the research questions underlying the studies of the thesis. The four studies were designed in a similar fashion, which is summarized in the method sections (Chapter 3). Following this, the studies are presented (Chapters 4 to 7). Finally, findings are summarized (Chapter 8) and discussed on a comprehensive level with an emphasis on issues and results that were found in all studies (Chapter 9).

Chapter 2

Training in process control

This chapter leads to the research question that guided the four studies presented in this doctoral thesis and introduces the model underlying the studies. The studies are integrated into present theory and research, and their similarities are highlighted.

2.1 Characteristics of process control

Despite growing concerns relating to nuclear technology after Three Mile Island (1979) and Chernobyl (1986), there are now 438 nuclear power reactors operating worldwide and the number is growing. More than 40 nuclear power plants are under construction around the world, and 80 more are planned. Reactors are predominantly built in Asia. Ten reactors are under construction in China, seven in Russia and six in India (Gillmann, 2009; IAEA, 2009; Meshkati, 1991; Vicente, 2006). The nuclear industry is one example for process control. However, the task of process control can be found in a range of industries that regulate and control large chemical, energy or thermal processes such as refineries, chemical plants or steel making (Moray, 1997; Wickens & Hollands, 2000). In such industries, the safety and reliability of plants is a critical goal and the ongoing construction of new plants emphasizes the need for continuing efforts relating to safety and reliability.

Process industries involve processing of materials and energy to produce a product by means of physical or chemical transformations (Moray, 1997). The processes are highly safety-critical and disturbances or accidents can have severe consequences for nature and for humans (Wickens & Hollands, 2000). A great deal of research effort has been undertaken to analyze the relation between errors and accidents, to find out what kind of errors resulted in accidents, and in particular what proportion of errors was caused by humans (directly) and what proportion was caused by technology failures (e.g. Reason & Hobbs, 2003). In human-machine interaction, 60 to 90% of all system failures are attributed to human error, regardless of domain (Hollnagel, 1998). Thus, the human is a critical component to establish and ensure safety in man-machine systems.

Wickens and Hollands (2000) describe three particular characteristics of process control.

- The processes are generally highly complex and involve a high number of interacting variables and many degrees of freedom (Moray, 1997; Wickens & Hollands, 2000). Variables can be cross-coupled, so that changes in one variable affect several other variables simultaneously. Modern control rooms comprise more than 5000 displays and thousands of controls and alarms to display these processes (Sheridan, 2006; Vicente, Mumaw & Roth, 2004). Such complexity can severely overburden the operator's mental model of the status of the plant and makes it extremely difficult for operators to identify the state of the plant. The mental model of the status of the plant, however, is critical for both normal control and abnormal situations (Moray, 1997; Wickens & Hollands, 2000).

– The process variables and system responses are slow. A control action may not produce a visible system response for seconds or minutes. Because of these slow system responses, human performance regarding decision making, attention, perception or memory are essential for control while motor abilities are less important. Due to the slow responses, manual control of operators is mostly open-loop. The operator controls outer-loop variables, whereas automated adjustment and feedback loops handle the inner-loop control, as illustrated by the following example: "The operator of a blast furnace may choose a set point of desired temperature, and automated inner-loop control will provide the amount of fuel and energy to the furnace necessary to achieve that temperature some minutes later" (Wickens & Hollands, 2000, p. 514). A closed-loop strategy would be potentially inefficient and unstable because of the slow system responses.

– Process control is closely tied to automation (Wickens & Hollands, 2000; Sheridan, 2002). Many components of process control have been automated because of the immense complexity and the demands posed on the human operator. Toxic or hazardous materials, for example, cannot be handled by the operator directly. On the one hand, automation supports and relieves the operator, but on the other, automation can entail difficulties such as complacency, decreased situation awareness or loss of skills (Manzey, 2008). Bainbridge (1983) points out the irony that one does not necessarily remove difficulties by automating. However, process control is not synonymous with automation, because many processes can be controlled manually. Typical situations for manual intervention include the start-up and shut-down of the plant, abnormal situations and fault management (Moray, 1997).

2.2 Training objectives in process control

Training of personnel is one of the measures from among other safety-supporting activities such as system and interface design to assure safe operation in process control. Regular training of operators takes on an important role in process industries. The objective of training is to minimize errors and to prepare personnel to cope with incidents, abnormal situations, and even the worst case scenario (Flexman & Stark, 1987; IAEA, 2004; Mannarelli, Roberts & Bea, 1996; Wickens & Hollands, 2000).

The starting point for training design is the determination of training objectives. Therefore, the first and crucial step in training development includes the analysis of the task and task components to be trained. Based on a cognitive task analysis including different methods such as error analysis, hierarchical task analysis and protocol analysis, cognitive requirements of the process control task were derived in a pilot study (Burkolter, Kluge, Schüler, Sauer & Ritzmann, 2007).

The operator's task in process control has been described by Wickens and Hollands (2000) as "hours of intolerable boredom punctuated by a few minutes of pure hell" (p. 517). Although it does not apply exactly to the work in process industries and is somewhat overstated in this case, the assertion illustrates the two major tasks in process control: The first task of the operator is to control and stabilize the system by standard procedures, which corresponds to the "hours of intolerable boredom" (Wickens & Hollands, 2000). The second task involves intervening, diagnosing and repair in the case of fault states and abnormal situations, i.e. the "minutes of hell" (Wickens & Hollands, 2000). The two tasks, system control and fault finding, demand different cognitive regulatory processes.

System control mainly requires trained procedures to be followed, which mostly occurs on a rule-based level of cognitive control. In contrast, fault finding during novel situations may require innovation because no previous know-how or rules are available. Thus, the control moves on a higher level, the knowledge-based level of cognitive control (Rasmussen, 1990; Wickens & Hollands, 2000). For system control, the operator is focusing on the forward flow of events ("what causes what?"), whereas the operator needs to reverse this thought pattern during fault states ("what was caused by what?"). Finally, it is important to note that ability in system control and fault finding are independent. An operator who performs well in system control and stabilization need not necessarily be as good in fault finding, and vice versa (Landeweerd, 1979; Wickens & Hollands, 2000).

Abnormal situations and fault states (fortunately) occur very rarely in practice. However, operators have to be prepared for these cases, because consequences of errors can be grave. Moreover, there are procedures such as start-up and shut-down of plants that only require implementation at certain intervals. Thus, operators need to retain knowledge and skills for long durations, despite periods of non-use. Those periods of non-use can last for several months or years (called temporal transfer). A further challenge is that knowledge and skills might have to be adapted to novel situations because it is not possible to prepare for every fault situation beforehand (called adaptive transfer; Kluge, Sauer, Schüler & Burkolter, 2009; Sauer, Hockey & Wastell, 2000b).

Taken together, training objectives for process control can be summarized as follows: The operator has to be trained to perform the two main tasks of process control, system control and fault finding, both for practiced and novel fault states.

2.3 Models of training antecedents and training outcomes

The extent to which training objectives are reached – system control and fault finding for practiced and novel fault states – is evaluated by measuring training outcomes. Training outcomes are influenced by training antecedents such as individual and situational characteristics or different training methods. These relationships between training antecedents, training methods and training outcomes are analyzed in the present studies. A model of training antecedents, training methods and training outcomes was developed as a basis for the studies of this thesis.

The model was derived from (a) the integrative theory of training antecedents and training outcomes by Colquitt, LePine and Noe (2000) and (b) the classification scheme for evaluating training outcomes by Kraiger, Ford and Salas (1993). Both the theory and the classification scheme are described in more detail in the next sections. Additionally, reasons for basing the model on the two approaches are given.

2.3.1 Integrative theory of training antecedents and training outcomes

Colquitt et al. (2000) derived their integrative theory of training antecedents and training outcomes based on a literature review and meta-analysis of studies from work and organizational psychology. The meta-analysis included more than a hundred studies on training conducted in the field (business organizations and military) and the laboratory. In their theory, training antecedents are linked to training outcomes (see Figure 2.1).

Several reasons are put forward as to why the integrative theory of training antecedents and training outcomes was chosen upon which to base the

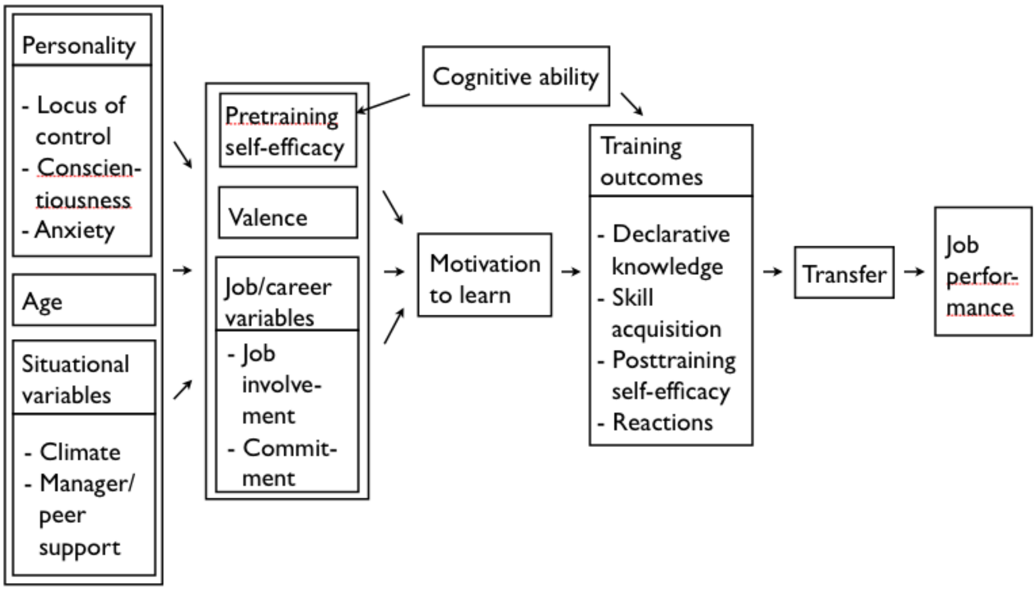

Figure 2.1: Colquitt and colleagues' (2000) integrative theory of training motivation, its antecedents, and its relationships with training outcomes (simplified)

studies of the thesis. As explained in the introduction, this thesis aims at integrating research approaches and findings from work and organizational psychology as well as human factors with the aim of gaining new insights. Since a comprehensive and suitable training theory for process control tasks cannot be found to date, a training theory from work and organizational psychology was chosen to apply to process control. The model by Colquitt et al. (2000) is found suitable for process control, because it includes the components that are relevant for training in process control. The theory presents and describes factors affecting training outcomes. Furthermore, the theory also includes transfer performance as an additional factor relating to training outcomes. As described in section 2.2, temporal and adaptive transfer are critical goals in the training of process control tasks. Finally, Colquitt et al. (2000) provide a model that is not only theoretically based, but also empirically tested.

Training antecedents include individual and situational characteristics, pretraining self-efficacy, valence, job/career variables, cognitive ability and motivation to learn. Individual characteristics (e.g. locus of control or conscientiousness) as well as situational characteristics (e.g. climate, manager/peer support) have an effect on training outcomes. Training outcomes, in turn, influence transfer and job performance. The effect of individual and situational characteristics on training outcomes is mediated by pretraining self-efficacy, valence and job/career variables and motivation to learn. Furthermore, it was shown that individual and situational characteristics are also directly related to training motivation, training outcomes, transfer and job performance. Because the model proposes both mediated as well as direct relationships, Colquitt et al. (2000) refer to it as the partially mediated training theory.[1] They suggest that individual and situational characteristics are critical factors before training, during training, and after training.

2.3.2 Classification scheme for evaluating training outcomes

Colquitt et al. (2000) chose declarative knowledge, skill acquisition, post-training self-efficacy and reactions as measures of training outcome. However, critical measures of training outcomes for process control such as procedural or structural knowledge are lacking in this approach. Therefore, the measurement of further training outcomes connected with a structural approach to the measurement of training outcomes is suggested. This approach is described below.

[1]Colquitt et al. (2000) not only present the mediated relationship between individual and situational characteristics and training outcomes, but also the direct relationship between individual and situational characteristics and each of the factors motivation to learn, training outcomes, transfer and job performance. For the sake of clarity, these direct relationships are not depicted in Figure 2.1.

Assuming that learning outcomes are multidimensional, Kraiger et al. (1993) propose a novel classification scheme for evaluating learning outcomes. By multidimensional, Kraiger et al. (1993) mean that "learning may be evident from changes in cognitive, affective, or skill capacities" (p. 311). In earlier approaches of training evaluation (e.g. Kirckpatrick, 2007), it was not clear how training was conceptualized or how learning should be measured (Kraiger et al., 1993). Therefore, drawing from research in cognitive, social and instructional psychology as well as human factors, three categories –cognitive, skill-based and affective training outcomes– are suggested. A short description of the three training outcomes and their constructs following Kraiger et al. (1993) is given below.

Cognitive outcomes It is questioned whether measures of verbal knowledge are capable of discriminating "among learners at higher levels of cognitive development" (Kraiger et al., 1993, p. 313). Therefore, Kraiger et al. propose structural knowledge and cognitive strategies as cognitive outcomes in addition to verbal knowledge. Structural knowledge is perceived as the way in which concepts are organized and interrelated within a knowledge domain (Davis, Curtis & Tschetter, 2003).

Skill-based outcomes In the initial skill acquisition phase, trainees' performance is slow, relying strongly on working memory. In comparison, advanced skills are characterized by smooth and fast performance. With continued training, automaticity can be reached. The automatic processing allows the trainee to cope with additional demands because greater cognitive resources are available.

Affective outcomes Affective outcomes can influence behavior and performance. In short, all learning outcomes that are neither cognitive nor skill-based are counted as affective outcomes. Affective outcomes include attitudinal or motivational outcomes. An attitudinal outcome might involve changing values, while motivational outcomes are often secondary objectives of training, for example self-efficacy.

2.4 Model of training antecedents, training methods and training outcomes in process control

From the theory and the classification described, a model of training antecedents, training methods, training outcomes and transfer performance for the context of process control was derived (see Figure 2.2). The model depicts how training antecedents affect training, and training in turn affects training outcomes as well as transfer performance. It was developed in line with the theory and the relationships between training antecedents and outcomes proposed by Colquitt et al. (2000) and was complemented with the classification scheme for training outcomes by Kraiger et al. (1993). However, the model was adapted to suit the characteristics of the process control task and its training, as presented above (see sections 2.1 and 2.2).

The research questions of this thesis are based on the model. The guiding idea of this thesis is to support operators with training in consideration of operator characteristics. The research question of this thesis is as follows: How can training best support process control performance with regard to operator characteristics? And more specifically: Which operator characteristics, training methods and which knowledge type support which training outcomes in process control best? To approach the research question, the components of the model and their relationships with each other are analyzed in the studies.

The model by Colquitt et al. (2000) lacks to mention the delivery of training and its effects on training outcomes explicitly. Different training methods

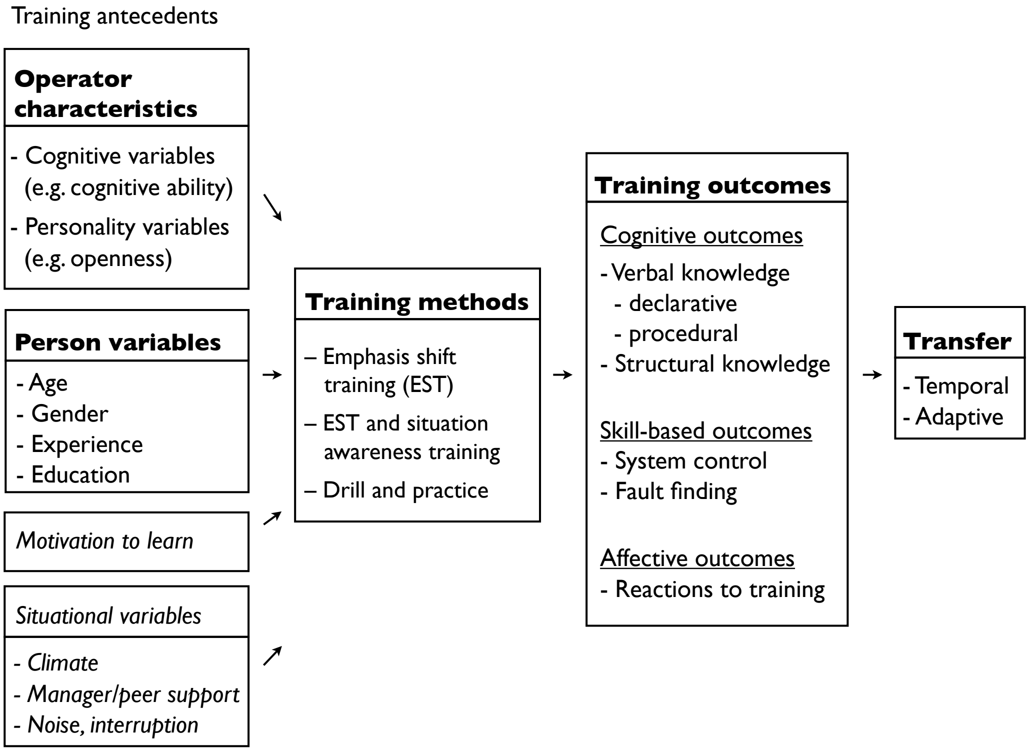

Figure 2.2: Central model underlying this thesis: Effects of training antecedents and training methods on training outcomes and transfer (derived from Colquitt et al., 2000 and Kraiger et al., 1993). Measures printed in bold letters are reported in the studies.

may have different effects on training outcomes. Therefore, the delivery of training with different training methods has been added as a component in the model of this thesis. Training was an integral part of all studies and the participants were trained with approaches especially designed for process control. In Study III, the influence of different training methods on training outcomes was specifically investigated. The study and the training methods emphasis shift training (EST), EST and situation awareness training, and drill and practice will be described in more detail in section 2.4.2.

While training methods have been added as a component to the model of this thesis, some components as suggested by Colquitt et al. (2000) needed to be excluded. These exclusions concerned training antecedents. Training antecedents relating to individual and situational characteristics have been included in the studies, while valence and job/career variables have not been assessed. Moreover, on-the-job performance was not measured. These variables that are relevant in an organizational context such as job/career variables (e.g. organizational commitment) and job performance in an organization were not focused upon. The focus of this thesis was on the individual level, i.e. training of the individual operator. However, the group-, organizational- and political levels are also of importance for safe processes (cf. Rasmussen, 1997). For instance, process industries usually involve several operators working as an integrated crew (Moray, 1997), emphasizing the relevance of the group level.

As operator characteristics, both cognitive (e.g. cognitive ability, working memory capacity, decision making) and personality variables (e.g. openness, conscientiousness, pretraining self-efficacy) and their effect on training outcomes were assessed. Person variables such as age and gender were collected along with the experiments, primarily to control for comparableness of experimental groups. Motivation to learn and situational variables, however, were not explicitly measured or reported in the studies (and therefore printed in italics in Figure 2.2). Moreover, the studies of this thesis were not designed to analyze mediated effects as in the meta-analysis of Colquitt et al. (2000, with meta-analytic path analyses). Therefore, motivation to learn, cognitive ability or pretraining self-efficacy are not depicted as mediators in Figure 2.2.

Following Colquitt et al. (2000), performance was assessed by measuring training outcomes and transfer. Training outcomes were structured as suggested by Kraiger et al. (1993) and hence include cognitive, skill-based and affective outcomes. Cognitive outcomes include declarative, procedural and structural knowledge. Skill-based outcomes include performance on the two main tasks of process control-system control and fault finding. Lastly, reactions to training were assessed as affective outcomes. Cognitive and skill-based outcomes were analyzed in all studies while affective outcomes were only investigated in one study (Study III). Structural knowledge as a cognitive outcome was a special focus of Study IV.

All four studies of this thesis have in common that they investigate specific *effects on training outcomes*. While Studies I and II investigated the effects of operator characteristics on training outcomes, Studies III and IV investigated the effects of training on training outcomes. In Study III, the focus was on the comparison of different training methods to enhance process control performance, while Study IV put an emphasis on methodological issues, assessing a novel measure as a training outcome for process control.

2.4.1 Effects of operator characteristics on training outcomes

Although Study I investigated the relationship between operator characteristics and process control performance in the context of personnel selection, operator characteristics are essential for training, too. As depicted in the model underlying this thesis, individual characteristics of trainees, that is, characteristics that trainees bring with them to a training program, can influence training outcomes. For instance, high-ability trainees are more likely to learn and succeed in training when all other things are equal (Salas et al., 2006b). The importance of individual differences in the ability to learn has long been a theme in educational psychology and instructional psychology, while it seems to be less prevelant in the context of human factors and ergonomics. Learners transform what they receive from training and instruction and construct their knowledge by themselves. Therefore, what the learner brings to the training situation is of crucial importance (Pintrich, Cross, Kozma & McKeachie, 1986). Mumford, Harding, Weeks and Fleishman (1988) conducted a comprehensive study to examine the contributions of individual characteristics and course-content variables on training effectiveness. They found that intellectual variables (as individual characteristics) had a greater impact on achievement than course-content variables such as instructor experience or instructional aids.

Several reasons can be put forward as to why understanding and knowledge about operator characteristics and their relationship with performance is relevant. First, an understanding of operator characteristics can be helpful in designing training appropriately, by customizing training to the needs of trainees. Ideally, training is designed to accommodate individual differences between trainees (Patrick, 1992). Second, knowledge about operator characteristics can provide the basis for decisions in personnel selection and the development of selection tools. Besides application in training and selection, Meyer, Nachtwei and Kain (2009) as well as Szalma (2009) argue for the general incorporation of an individual differences approach into human factors research and practice: "Description of the systematic variation in the human portion (e.g. cognitive and personality traits; motivational and emotional states) of human-technology systems can complement the existing design methods (e.g. task analysis) to yield better models of system performance and improve system design and operation" (Szalma, 2009, p. 381). Thus, research on operator characteristics can also be seen as supplementing task analyses to provide a better understanding of process control and its requirements. Moreover, in accordance with Szalma and Meyer et al., the view is taken that analysis of individual characteristics should be an integral part of human factors studies in general.

So far, reviews and meta-analyses (Colquitt et al., 2000; Salas et al., 2006b) have shown that individual characteristics such as cognitive ability, self-efficacy, conscientiousness, locus of control, anxiety, valence, goal orientation and motivation affected training outcomes in various domains. However, the issue of individual characteristics and their influence on performance has rarely been raised in process control. Therefore, in Studies I and II, research on individual characteristics was conducted by analyzing individual characteristics with a possible connection to process control performance.

Study I investigated the relationship between operator characteristics and process control performance. Study II aimed at confirming and extending findings of Study I by analyzing the relation between the same and additional operator characteristics and process control performance on the basis of two other experiments. Working memory capacity, set-shifting performance and decision making were analyzed in addition to cognitive ability and cognitive flexibility, which were assessed as cognitive variables in Study I. Set-shifting performance is perceived as the ability to establish and then shift responses or tasks (Nagano-Saito, Leyton, Monchi, Goldberg & He, 2008). Need for cognition (Cacioppo & Petty, 1982) and perfectionism (Altstötter-Gleich & Bergemann, 2006) were assessed in addition to personality traits and self-efficacy, which were analyzed in Study I. Need for cognition describes the tendency to engage in and enjoy thinking (Cacioppo & Petty, 1982).

2.4.2 Effects of training methods on training outcomes

There is a long tradition of investigating effects of different training methods on training outcomes (Quiñones & Ehrenstein, 1997). A range of training methods have been reviewed regarding their effectiveness in the context of human factors and complex work environments, for instance by Morris and Rouse (1985), Salas et al. (2006b) or Kluge et al. (2009). The reviews all have in common that they refer –amongst other training methods– to technology-oriented training methods such as simulation-based training and games as well as simulator-based training. Salas et al. (2006b) describe technology-oriented training methods as methods that use "technology to provide opportunities for practice and instruction in realistic settings" (p. 484). Simulation-based and simulator-based training is also regularly employed in process control and other high-risk industries (Kluge et al., 2009; Mannarelli et al., 1996; Wickens & Hollands, 2000). Simulation-based training provides possibilities for training that other approaches cannot, such as stopping a simulation during training. In Study III, this advantage of simulation-based training methods was made use of (cf. Saus et al., 2006).

In Study III, three training methods were designed to enhance process control performance. The objective was to support learners by providing them with attention management strategies in order to reduce their mental workload. Attention capacities are crucial to process control performance. However, as they are limited, process control operators must learn to apportion their attention strategically (Gopher, 1996). For novice learners especially, highly complex tasks such as process control are very demanding (Gopher, Weil & Siegel, 1989). Therefore, the goal in Study III was to support novice learners by providing them with attention management strategies. All three training methods sought to improve attention management skills and ultimately performance.

The first training approach designed to enhance attention management skills and process control performance was emphasis shift training (EST; c.f. Gopher et al., 1989). In EST, multiple priority changes on subcomponents are given, while the whole task is left intact and only the attention level is changed (Gopher et al., 1989). The underlying idea is to provide learners with strategies to cope with complexity and thereby reduce mental workload and improve performance. The second training approach combined EST with a situation awareness training (EST/SA). Situation awareness (SA) is critical to process control performance (Endsley, 1995; Wickens & McCarley, 2008). To improve individuals' SA, training of attention management is recommended (Endsley & Robertson, 2000). By combining EST with SA training, an approach to enhance SA and performance directly (with the SA training) and indirectly by training attention and task management to support SA (with the EST) was chosen. SA was practiced through the freezing technique with debriefing (as described by Saus et al., 2006). The simulation was randomly stopped, and SA questions were then posed and answers debriefed. Finally, with drill and practice (D&P), a trainee learns a task by means of repetition and rehearsal until some level of proficiency is reached (Cannon-Bowers, Rhodenizer, Salas & Bowers, 1998). Practice can support attention performance by proceduralizing or automating a task in order to free up resources for another task. Continuous practice in a task was shown to lead to improved attention skills (Wickens & McCarley, 2008).

While Study III aimed at comparing the effects of different training methods on training outcomes, Study IV aimed at investigating *how effects of training methods on training outcomes can be measured*. Thus, in Study IV, a methodological approach to the analysis of effects on training outcomes was taken.

Aguinis and Kraiger (2009) noted in their review on training that there have been several conceptual contributions for training evaluations in recent years. However, there has been little empirical work on validating new training outcome measures. Study IV aimed to contribute to closing this research gap by evaluating a novel measure as a training outcome in process control. Based on the classification scheme by Kraiger et al. (1993), Study IV aimed to evaluate the potential of a novel measure for structural knowledge as a training outcome in process control and to achieve a better understanding of declarative, procedural and structural knowledge of operators. Traditionally, knowledge is assessed by verbal achievement tests on the subject matter. Traditional methods, however, are regarded as limited in their ability to assess higher-order learning (Kraiger, Salas & Cannon-Bowers, 1995). Therefore, a novel measure of structural knowledge was assessed in addition to verbal knowledge. The Association Structure Test (Meyer, 2008) integrates an association task and pathfinder network scaling based on relatedness ratings into one computerized testing system. The Association Structure Test was employed in two experiments together with verbal tests on declarative and procedural knowledge in order to assess whether incremental variance in performance could be explained by the novel structural knowledge measure.

Chapter 3

Method

In the following sections, the simulated work environment that was employed in all studies will be presented. Thereafter, the design, samples, and measures of the studies will be summarized.

3.1 Simulated process control environment

The four studies have in common that they all employed the same simulation of a process control task. Using the same experimental task ensured that findings could be related to each other.

3.1.1 Theoretical approach to simulation design

The studies of this thesis were conducted with the computer-based process control simulation called Cabin Air Management System (CAMS, version 3.0; Sauer, Wastell & Hockey, 2000a, see Figure 3.1). CAMS was developed as part of a research program of the European Space Agency following a theoretical approach for micro-world design.

Two key issues were essential to the framework. First, a human factors analysis of spaceflight including a task analysis, an expert consultation and a comparison of analog domains was conducted to identify relevant tasks in the work environment. Management of the life support system was chosen

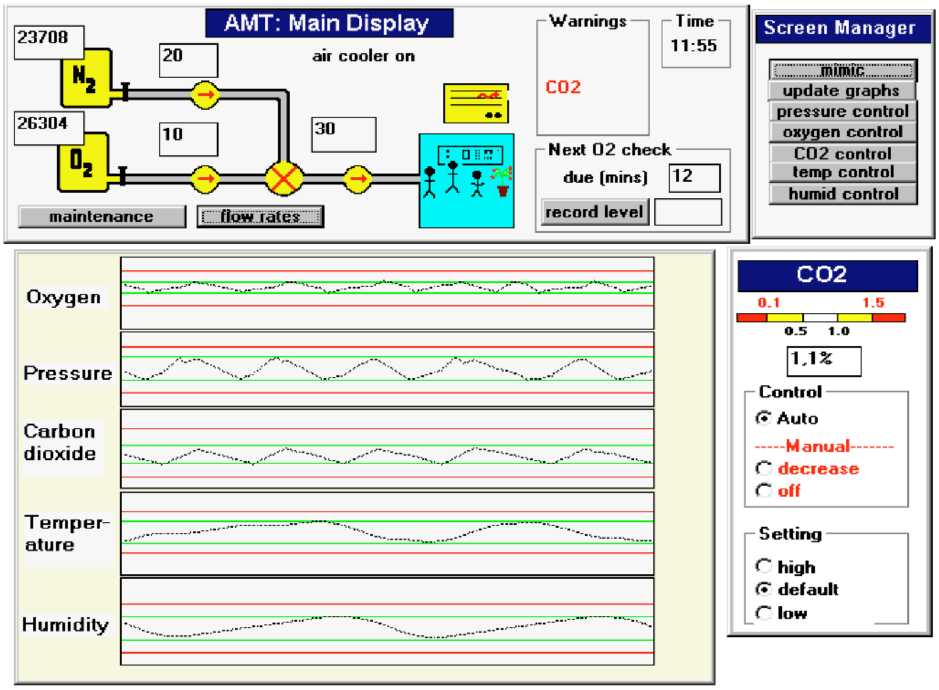

Figure 3.1: Main display of CAMS (Sauer, Wastell & Hockey, 2000) with history display, warning system, screen manager and CO2 control panel

as a safety-critical subset of activities for CAMS. The life support system essentially represents a process control system. Thus, although CAMS is placed in the context of spaceflight, its underlying principles correspond to a process control task (Sauer et al., 2000a).

The second key issue central to the design framework was the use of a theory of human performance – the compensatory control model of performance regulation by Hockey (1997). The theory suggests that changes in work demands lead to adaptive human behavior. For instance, CAMS includes tasks that have different priorities attached, because the theory proposes that performance decrements are more likely in low-priority tasks than in high-priority tasks.

There are several reasons why CAMS was chosen as an experimental task: First, CAMS clearly represents a process control task characterized by high complexity, opaqueness, dynamics and time-lags. The main tasks of process control – system control and fault finding – are represented. Second, CAMS provides the possibility to assess a trainee's workload by analyzing performance on the secondary tasks. Third, for training and testing of temporal and adaptive transfer of fault finding, CAMS allows for the programming of a range of system faults, enough to use some in training and novel ones in the testing session. Finally, CAMS has been successfully employed in a number of previous studies, including training experiments (e.g. Sauer et al., 2000b; Hockey, Sauer & Wastell, 2007).

The use of simulations or microworlds represents "a compromise between experimental control and realism" (Gonzalez, Vanyukov & Martin, 2005, p. 274). On the one hand, simulations enable controlled research to a higher extent than would be possible in a real-world setting. On the other hand, the external validity can be questioned, as the simulated environment does not necessarily represent real-world tasks or requirements. However, it is assumed that thoughtfully designed simulations represent the essential demands of the work environment (cf. Gonzalez et al., 2005). In the case of CAMS, the validity was supported by the theoretically based design process (cf. Sauer, Burkolter, Kluge, Ritzmann & Schüler, 2008).

3.1.2 The tasks of the operator

CAMS integrates the two major tasks of process control: System control and stabilization on the one hand and detection, diagnosis and repair of system faults on the other hand. Moreover, CAMS simulates a number of characteristics of process control: A high level of automation, opaqueness, time-lags, closely coupled subsystems, restricted access to system controls, false alarms as well as a dynamic autonomous underlying process.

The simulated process control environment consists of five main parameters (O_2, CO_2, cabin pressure, temperature, humidity) which are kept within their target zone by automatic controllers. If one or more parameters depart from the target zone, the operator has to intervene. System control and fault finding are defined as primary tasks, while alarm acknowledgement and tank level recordings are seen as secondary tasks (Sauer et al., 2000a). The four tasks of the operator are described below.

System control The operator needs to monitor the system closely to detect deviations of parameters from the target zone. In the case of a deviation, one can intervene by adjustment of automatic control parameters or manual control.

Fault finding Eighteen system faults can be programmed into CAMS. In case of a system fault, the cause has to be diagnosed with suitable tests and repaired by means of the maintenance facility.

Tank level recordings (prospective memory task) O_2 tank levels have to be recorded every third minute without further prompt. This task corresponds to a prospective memory task.

Alarm acknowledgement (reaction time task) Appearing warnings have to be confirmed by clicking on them. False alarms can be detected by looking at parameter levels. Alarm acknowledgement is basically a reaction time task.

3.2 Design of the studies

An overview of the samples as well as the dependent and independent variables of the four studies can be found in Table 3.1.

Table 3.1: Overview of samples, dependent and independent variables of all four studies

	Sample	Independent Variables	Dependent Variables
Study I	39 trainee operators	- Cognitive variables - Personality variables	- System control - Fault finding - PMF - Reaction time - Declarative kn.
Study II	41 engineering students	- Cognitive variables - Personality variables	- System control - Fault finding - Declarative kn.
	50 engineering students	- Cognitive variables - Personality variables	- System control - Fault finding - Declarative kn.
Study III	40 engineering students	- Training method - Time of measurement - Practiced/novel faults	- System control - Fault finding
Study IV	41 engineering students	- Declarative knowledge - Procedural knowledge - Structural knowledge	- System control - Fault finding
	50 engineering students	- Declarative knowledge - Procedural knowledge - Structural knowledge	- System control - Fault finding

Notes: The four studies are based on data of three training experiments (N = 39/41/50). PMF = Prospective memory failures; kn. = knowledge.

3.2.1 Design

The research questions in this thesis were all approached using experimental studies. The experiments were all similarly designed and included between-subjects as well as within-subject factors. Multifactorial designs with two or three points of measurement were employed. In all studies, there was at least one week between training and testing sessions, but up to six weeks.

This was essential as transfer performance was to be measured in all of the studies.

3.2.2 Participants

In all studies, trainee operators working in chemical plants or engineering students were invited as participants. This was to ensure that participants had some knowledge and understanding of technical systems. The mean age of the participants ranged from 18 to 25 years. Hence, young adults at the beginning of their vocational careers took part in the experiments. In all studies, samples included both men and women. In summary, the samples of the studies are comparable regarding age, sex and education.

3.2.3 Measures

Measures are deduced from the model of training antecedents, training methods and training outcomes underlying this thesis. Dependent variables were basically the training outcomes, i.e. cognitive, skill-based and affective training outcomes were measured as dependent variables. In Study IV, however, knowledge measures were assessed as independent variables, in contrast to the other studies. Since the objective of Study IV was to assess a novel method for structural knowledge assessment, the relationship between knowledge measures and skill-based training outcomes was analyzed.

Following the description of the theoretical and empirical background, the research questions and the method used in the studies, the four studies are presented in the next chapters.

Chapter 4

Study I: Predictive qualities of operator characteristics

Burkolter, D., Kluge, A., Sauer, J. & Ritzmann, S. (2009). Predictive qualities of operator characteristics for process control performance: The influence of personality and cognitive variables. *Ergonomics, 52*, 302–311.

Ergonomics
Vol. 52, No. 3, March 2009, 302–311

Taylor & Francis
Taylor & Francis Group

The predictive qualities of operator characteristics for process control performance: The influence of personality and cognitive variables

Dina Burkolter[a]*, Annette Kluge[a], Jürgen Sauer[b] and Sandrina Ritzmann[a]

[a]*Research Institute for Organizational Psychology, University of St. Gallen, Varnbüelstrasse 19, CH-9000 St. Gallen, Switzerland;* [b]*Department of Psychology, University of Fribourg, Rue de Faucigny 2, CH-1700 Fribourg, Switzerland*

This article examines the relationship between operator characteristics and process control performance. Thirty-nine trainee operators participated in a 4-h training session of a simulated process control task and a testing session in which various system faults had to be managed. Cognitive ability, cognitive flexibility, self-efficacy and personality traits were measured as operator characteristics. Cognitive ability related positively to system control performance but not to diagnostic performance. Participants with low cognitive flexibility performed best on system control, whereas participants with high cognitive flexibility performed best on diagnostic performance. A hierarchical regression analysis revealed that cognitive ability, cognitive flexibility and declarative knowledge accounted for about 30% of the variability of system control. The findings suggest that consideration of cognitive ability and cognitive flexibility be increased in personnel selection for complex work environments.

Keywords: process control; individual characteristics; personnel selection; performance

1. Introduction

Safety, reliability and productivity have long been key objectives in process control environments such as refineries and nuclear power plants. They remain high priorities for process industries because human errors and their consequences may cause severe damage for nature and human beings (Wickens and Hollands 2000). How can these goals be achieved? There are different ways in which organisations strive for high reliability and safe production. One typical approach in the field of ergonomics is to increase the emphasis on system or task design, that is, to improve systems or interfaces in order to enhance operators' performance. The focus is thus on improving the machine side of the human–machine system. However, highly reliable and safe production cannot be achieved by tending only to the complex system and its components; it depends on qualified and competent personnel as well (International Atomic Energy Agency 1996). Enhancing the qualification of personnel through regular training is therefore an additional approach taken by organisations. A third way to contribute to safety, reliability and productivity in process control industries is to select personnel, such as process control operators, who are most likely to succeed at their jobs. This raises the question of which operator characteristics can best predict process control performance.

There is a small branch of ergonomic research that addresses issues of individual characteristics, performance and personnel selection. Hollnagel (1998), for instance, differentiates between temporary person-related variables (e.g. inattention and fatigue) and permanent person-related variables (e.g. variables relating to cognition), both of which are linked to human errors. Wickens *et al.* (1998) address the matter of selection in human factors engineering, drawing on many empirical findings from aviation research, a field in which this subject has long figured prominently (Hunter and Burke 1994). Coverage of empirical findings on selection in process control environments is harder to come by.

Three reasons are advanced to explain the lack of research on the selection of process control operators (Williams and Taylor 1993, Stanton and Ashleigh 1996). First, research on operator selection is considered complex because of the methodological problems entailed by this type of study. For instance, the infrequency and irregularity of events in process control make it difficult to measure actual job performance. Second, human factors researchers might feel that selection has little effect on work performance in comparison to aspects such as system design and operator training. Third, research has shown that predictive power is quite poor (Stanton and Ashleigh 1996). However, Stanton and Ashleigh (1996) argue

*Corresponding author. Email: dina.burkolter@unisg.ch

ISSN 0014-0139 print/ISSN 1366-5847 online
© 2009 Taylor & Francis
DOI: 10.1080/00140130802376067
http://www.informaworld.com

that this assessment becomes a 'circular argument: because it is a difficult area to study and people perceive that it is less important than other areas, little research is done, so the predictive power of the methods is not optimised' (p. 181). The aim of this article is to improve understanding of how operator characteristics such as cognitive ability and style, self-efficacy and personality relate to process control performance.

1.1. Selection of process control operators

Within the process control industries, the nuclear industry appears to have the most standardised procedures for personnel selection. Stating that the selection of suitably qualified personnel is essential for the safe and reliable operation of a nuclear power plant, the International Atomic Energy Agency (IAEA) provides recommendations and guidelines for the recruitment of plant personnel (e.g. International Atomic Energy Agency 2002, 2006). Recruitment of personnel is expected to follow approved procedures that include both the specification of selection criteria and objective testing to assess applicants' aptitudes for the job (International Atomic Energy Agency 2002). Aside from education, IAEA-recommended selection criteria are work background and experience, problem-solving ability, emotional stability, motivation, initiative, communication skills and attitudes towards safety. Attributes related to safety culture include 'a questioning attitude, a rigorous and prudent approach and communication and learning abilities' (International Atomic Energy Agency 2002, p. 5).

In a survey of selection methods that was carried out in 34 organisations in 13 countries of the nuclear industry, Stanton and Ashleigh (1996) showed that interviews and references were the most frequently employed methods, followed by personality and cognitive tests that were applied by more than 60% of the organisations. Tests are employed to predict performance and to ensure that applicants are sufficiently qualified. It is thereby assured that 'training resources are not wasted on candidates who are unfit for the job' (International Atomic Energy Agency 2006, p. 3). Personality characteristics such as honesty, integrity and attitudes are assessed as well and interest inventories are used. Different methods are chosen depending on which attributes are of primary interest. Written tests are administered to assess knowledge; oral tests, to assess knowledge and attitudes; and performance examinations, to assess skills and attitudes (International Atomic Energy Agency 2006). However, it remains unclear just how abilities, personality or other individual characteristics affect performance, whether all characteristics are equally important for predicting job success, which of them are essential and which are simply 'nice to have'.

1.2. Operator characteristics and their effect on performance

Individual characteristics that affect work performance in general include cognitive abilities and style, self-efficacy and personality traits (e.g. Morris and Rouse 1985, Colquitt *et al.* 2000). Each of these characteristics exhibited by the applicant or expected to be critical particularly for process control performance are addressed below. This review of individual characteristics thus rests on empirical research on complex systems and process control, but it is also complemented by research in related fields (e.g. aviation) because empirical research on this topic in the field of process control is rather scarce.

1.2.1. Cognitive ability

Across many work environments a strong relationship between cognitive ability and performance has been shown to exist indirectly through job knowledge (Schmidt and Hunter 1998). Regarding process control and related tasks, Morris and Rouse (1985) found in their review that cognitive ability was related to troubleshooting performance and that measures of job-related knowledge and skills had the highest relationships to performance.

A large number of interrelated variables account for a high level of complexity in process control (Wickens and Hollands 2000). Controlling complex systems therefore requires different aspects of cognitive ability. First, an operator needs reasoning ability in order to develop hypotheses about the causal structure of the system. Second, reasoning with numbers has proven to be a predictor of performance in dealing with simulated complex systems (Wittmann and Hattrup 2004, Kluge 2008). Third, an operator requires verbal and figural abilities in order to understand instructions, information and graphs given by the system (Wittmann and Hattrup 2004, Kluge 2008). Drawing on the empirical results presented in the literature, it is hypothesised that cognitive ability is positively related to process control performance.

1.2.2. Cognitive flexibility

Several measures of cognitive style such as field independence–dependence and reflectivity–impulsivity have been shown to be related to troubleshooting performance (Morris and Rouse 1985). Cognitive flexibility as a cognitive style is described by Spiro *et al.*

(1996) as a reductive and an expansive style of thinking. Individuals with a reductive cognitive style (i.e. low cognitive flexibility) have a preference for simplicity and rigid prescriptions from memory and an intolerance of ambiguity. Individuals with an expansive and flexible cognitive style (i.e. high cognitive flexibility) assume the world to be heterogeneous and have a flexible and situation-adaptive assembly of knowledge. Supported by empirical findings, Spiro *et al.* (1996) argue that individuals with high cognitive flexibility are more likely to perform required cognitive operations for knowledge acquisition in complex and ill-structured domains. Also, individuals with high cognitive flexibility are more qualified to learn in novel, complex situations. Following Spiro *et al.* (1996), it was expected that cognitive flexibility would be positively associated with the handling of complex systems.

1.2.3. Self-efficacy

A meta-analysis by Stajkovic and Luthans (1998) has shown self-efficacy to be correlated to work-related performance. As for complex problem solving, Debowski *et al.* (2001) found a positive relationship between self-efficacy and control performance. They stated that performance of complex activities is affected by self-efficacy in that individuals of high efficacy display more strategic flexibility, manage their time more efficiently and discard flawed strategies more quickly than others. Moreover, self-efficacy has been found to moderate the relationship between acquired knowledge and control performance (Kluge 2008). Thus, a positive relationship between self-efficacy and process control performance is assumed.

1.2.4. Personality

Personality traits such as conscientiousness, openness and emotional stability have been found to be related to work performance (e.g. Barrick and Mount 1991). In a meta-analysis by Barrick and Mount (1991), conscientiousness was related to performance for a range of occupational groups. Similar results have been found for pilots' performance, where conscientiousness was the strongest predictor (Pettitt and Dunlap 1995, cited Wickens *et al.* 1998, p. 559). Conscientiousness is believed to be particularly important for safety in process control, for many processes involve high risk and hazardous materials (Wickens and Hollands 2000). Concerning emotional stability, Bartram (1995) found that pilots who successfully completed the training programme were more emotionally stable than pilots who had not. Accordingly, it is believed that the personality traits of conscientiousness, openness and emotional stability correlate with process control performance.

1.3. Present study

The aim of this study was to examine the relationship between operator characteristics and performance in the field of process control. More precisely, cognitive ability, cognitive flexibility, self-efficacy and personality have been analysed, which previous studies have shown to be related to performance.

Operator performance as a criterion was measured in a simulated process control environment. The data-gathering facilities of the computer-based simulation made it possible to track various process control performance variables and measure both primary and secondary tasks. As primary tasks, system control and diagnostic performance were measured to take account of the independent character of the tasks. Since ability in system control is independent of ability in fault diagnosis (Landeweerd 1979), a competent controller may not necessarily be a competent diagnostician and vice versa (Wickens and Hollands 2000). As secondary tasks, regular recording of tank levels and acknowledgement of system alarms were also required of the operator. This setting consisting of primary and secondary tasks corresponded to a complex multitask environment.

The operator characteristics (cognitive ability, cognitive flexibility, self-efficacy and personality) that are expected to be predictors of process control performance were measured through various questionnaires.

A training session was conducted to instruct participants in the process control task. A testing session took place 1 week later. The issue of training effectiveness for process control environments was also examined in this study. The findings of this research question are reported in a separate article (Sauer *et al.* 2008). In the present article, the focus is on operator characteristics and its relation to performance.

2. Method

2.1. Participants

The study involved 39 participants (51.3% female) who were trainee operators at different Swiss chemical companies employing process control environments in their production systems. Participants with this background were sought for the sample to ensure that they had an understanding and knowledge of technical systems. As the trainees were given two half days off work to take part in the study it was not possible for the collaborating organisations to provide a larger sample of participants. Participants' ages ranged from 16 to 22 years (mean 18.1).

2.2. *Training*

Two different kinds of training were given. The participants in the control group received basic training after a procedure-centred approach, whereas the participants in the experimental group received additional heuristic rule training. The training and the findings on its effects are described in more detail in a separate article (Sauer *et al.* 2008). The training sessions lasted approximately 4 h in total. Training duration was limited to half a day, as it is current practice in some organisations (e.g. in chemical industries, Kluge *et al.* in press). The sessions were conducted with groups of nine or 10 participants working alone on individual computers. There were two training sessions based on procedure-centred approaches used in previous studies (e.g. Hockey *et al.* 2007). Heuristic rules were presented to one training group ($n = 19$), with the other group receiving a comparable task ($n = 20$).

Analyses were conducted on the whole sample. The authors did not expect any interaction effects of training method and operator characteristics on performance due to the considerable overlap of the training methods in their content. The two training approaches differed only in that in one group heuristic rules were added to the procedure-based training (totalling 25 min). Furthermore, there was no theory that would suggest an interaction of heuristic rule training and individual characteristics. However, to test that the assumptions were correct, the correlations between operator characteristics and performance were also statistically analysed as a function of training group, but the correlations revealed no relevant differences between the groups (Fisher's r-to-z transformation, all $p > 0.05$). Thus, training was not found to have a moderating effect on the relationship between operator characteristics and performance.

2.3. *Process control task*

Process control performance was assessed by using a computer-based simulation of a complex task called Cabin Air Management System (CAMS). CAMS has already been applied in a number of previous studies (e.g. Hockey *et al.* 2007). As the task has been described elsewhere (Sauer *et al.* 2000b), it will be outlined only briefly here. CAMS simulates a spacecraft's automated life-support system, but its underlying principles are analogous to a process control task. The operator must monitor the system

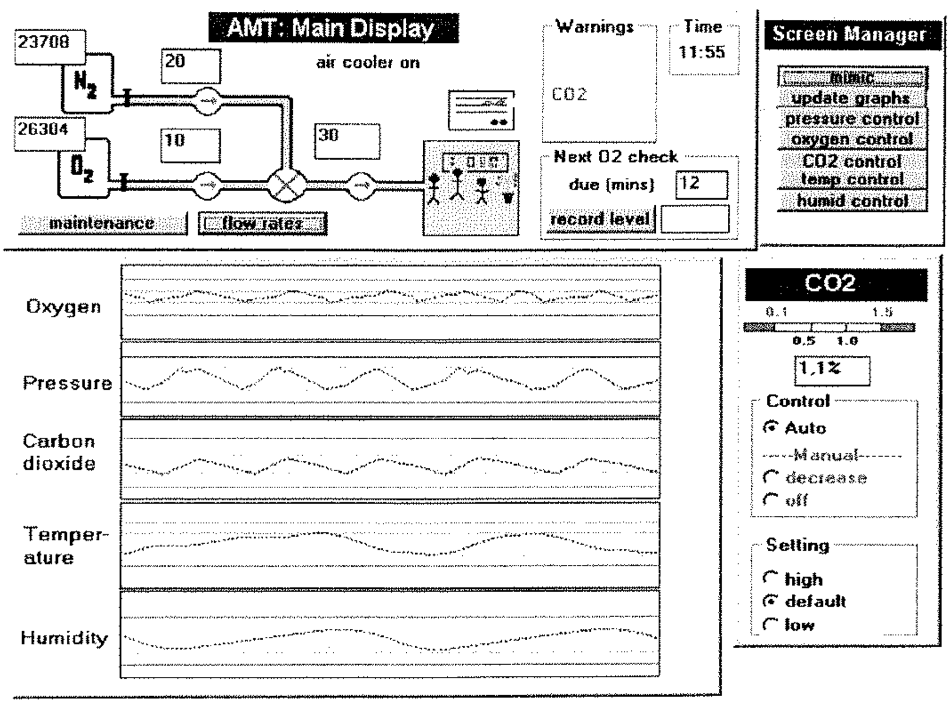

Figure 1. Main display of Cabin Air Management System.

and intervene if a system fault occurs. CAMS consists of five main system variables (O_2, CO_2, cabin pressure, temperature and humidity) that are maintained in a predefined zone by automatic controllers. The operator must complete four tasks with different priorities according to their importance for crew survival. Two of the tasks are defined as primary (system control and fault diagnosis) and two as secondary (prospective memory and reaction time; see below). Figure 1 shows the main display of CAMS with the screen manager, warning panel, history display, flow meters and its current values and control panels for the five main system variables.

2.4. *Assessment of operator characteristics*

Operator characteristics were assessed before the training session. The completion of the questionnaires took approximately 45 min.

2.4.1. *Cognitive abilities*

The Wonderlic Personnel Test (Wonderlic Inc. 2002) was used to measure cognitive abilities. The test consists of 50 items and captures verbal, numerical and figural aspects of intelligence and learning aptitude. The test questions included word comparisons, story problems and number series (e.g. 'What is the next number in this series? 1 0.5 0.25 0.125'). The participants had 12 min to work on the multiple-choice test.

2.4.2. *Cognitive flexibility*

Cognitive flexibility was measured with an adapted version of the Cognitive Flexibility Inventory (Spiro *et al.* 1996), which captures high and low cognitive flexibility. The scale consisted of three subscales that represent: (a) the preference to decompose complexity into parts vs. holism ('leaving complexity as it is'); (b) passive reception of information vs. active construction of information ('constructing knowledge'); (c) external vs. personal control of learning ('self-directed learning'). A 9-point scale ranging from -4 (one statement) to $+4$ (vs. other statement) was employed. A sample item from the scale was: 'Learning works best under the guidance of experts (e.g. teachers) vs. learning works best when it is self-directed'. Cronbach's alpha of the adapted German version was 0.70.

2.4.3. *Self-efficacy*

Self-efficacy was measured with eight items using Schyns and Collani's (2002) self-efficacy scale. A 6-point scale ranging from 0 to 100% (0%, 20%, 40%, 60%, 80%, 100%) was employed. Internal consistency of the adapted German version of the self-efficacy scale was satisfactory (Cronbach's alpha = 0.82) and similar to the original version (Cronbach's alpha = 0.88). A sample item from the scale was: 'When I am confronted with a problem in my job, I can usually find several solutions'.

2.4.4. *Personality traits*

Assessment of conscientiousness, openness and emotional stability was based on Saucier's (1994) Big-five Markers. Eight adjectives per personality trait (24 items total) had to be rated by participants on a 9-point scale ranging from 1 (extremely inaccurate) to 9 (extremely accurate). (Item examples for conscientiousness: organised, systematic, careless; for openness: creative, intellectual, complex; and for emotional stability: relaxed, moody, touchy.)

2.5. *Performance measures*

The data-gathering facilities of CAMS allow the measurement of performance criteria, including primary and secondary task performance. System control and fault diagnosis were defined as primary tasks; prospective memory and reaction time as secondary tasks (Sauer *et al.* 2000b).

2.5.1. *System control*

The operator must maintain five key parameters within a predefined zone. If a key parameter departs from the target zone, it can be restored to the proper level through adjustments of automatic controllers or adaptation of manual control. The duration of the parameters' deviation from the predefined zone was measured in seconds and converted into percentages.

2.5.2. *Fault diagnosis*

If a system fault occurs, the operator is required to diagnose and remedy it by means of the maintenance facility. One measure of diagnostic performance was the rate of faulty diagnoses; the other was the number of seconds the operator needed to remedy the system fault correctly.

2.5.3. *Prospective memory*

This secondary task involves the recording of O_2 tank levels every 3 min, a task corresponding to a prospective memory task (e.g. Brandimonte *et al.* 1996). Omitted tank-level recordings were counted.

2.5.4. *Reaction time*

The other secondary task is to acknowledge alarms when a warning signal appears. Reaction time is thereby measured. As false alarms can occur, the corresponding parameter must be examined after the acknowledgement.

2.5.5. *System knowledge*

Declarative knowledge (knowledge about the system and the relationship between the parameters) was assessed through a questionnaire. The latter has been employed before and is described in more detail in Sauer *et al.* (2000a). It consists of 12 multiple-choice items (e.g. 'What happens to pressure when the CO_2 scrubber is on?') with three alternatives ('increase', 'decrease', 'minimal or no effect'). Each set of alternatives was followed by an open question to explain the given answer ('Please explain why') and by three open questions about the processes and relationships in the CAMS environment (e.g. 'Please explain which components or processes have an impact on temperature in the cabin and describe the direction of the relationship'). Participants scored from 2.5 to 16.5 points of a possible 21 points (mean 10.0).

2.6. *Procedure*

The training session lasted approximately 5 h, including the completion of the questionnaires. The testing session lasted 2 h. Performance was measured in the testing session that was held 1 week after the training session in the same groups as the training. The participants worked on their individual computers for 70 min (two 35-min testing sessions), as they had done in their training. During the testing, they had to deal with three practised faults (faults they had trained to respond to), three novel faults (similar to practised faults, but not previously encountered) and two complex faults (which required an indirect way of handling the given fault) according to a schedule unknown to them. During the training session, the participants received a fault-finding guide, which they were also permitted to use in the testing session.

3. Results

3.1. *Correlations between operator characteristics and process control performance*

Data points more than 1.5 interquartile ranges away from the interquartile range were regarded as outliers. Two such outliers were identified for system control performance and were removed. One data record was missing because of technical problems. In this section, data of 36 participants are reported.

The overall pattern of the correlations between operator characteristics and process control performance revealed significant relationships between cognitive ability and performance and between cognitive flexibility and performance (see Table 1). Cognitive ability was significantly correlated with system control and system knowledge. Cognitive flexibility was significantly associated with both system control and diagnostic performance but, notably, in different directions. Participants with low cognitive flexibility generally performed better in system control than participants with high cognitive flexibility, whereas the latter participants tended to outperform the former in fault diagnosis and repair. There were no significant relationships between self-efficacy or personality and process control performance.

3.1.1. *Cognitive ability*

Cognitive ability was significantly correlated to system control failures ($r_s = -0.31$, $p < 0.05$). That is, individuals with high cognitive abilities showed better

Table 1. Spearman correlations between operator characteristics and performance.

Performance measure	Cognitive ability	Cognitive flexibility	Self-efficacy	Conscientiousness	Openness	Emotional stability
Primary task performance						
System control failures	−0.31*	0.34*	−0.18	−0.15	−0.04	−0.02
Diagnostic errors	−0.17	−0.31*	−0.24	−0.22	−0.14	0.00
Fault identification time	−0.19	−0.35*	−0.18	−0.13	−0.01	0.03
Secondary task performance						
Prospective memory failures	0.08	0.06	0.02	−0.09	0.23	−0.15
Reaction time	−0.13	−0.16	−0.01	−0.07	0.04	0.02
System knowledge	0.41**	0.06	0.15	−0.05	0.14	−0.27

*$p < 0.05$; **$p < 0.01$ (one-tailed).

system control performance than did participants with low cognitive ability. A significant relationship between cognitive ability and system knowledge was also found ($r_s = 0.41, p < 0.01$). In other words, the higher the cognitive abilities, the better the declarative knowledge.

There was no significant relationship between cognitive ability and overall diagnostic performance (diagnostic errors: $r_s = -0.17$, $p > 0.05$; fault identification time: $r_s = -0.19$, $p > 0.05$). However, an additional analysis for diagnostic performance of practised and novel faults showed that cognitive ability was significantly related to dealing with novel faults, both in diagnostic errors ($r_s = -0.30$, $p < 0.05$) and fault identification time ($r_s = -0.32$, $p < 0.05$). Cognitive ability was not significantly related to managing practised faults, though (diagnostic errors: $r_s = -0.07$, $p > 0.05$; fault identification time: $r_s = -0.01$, $p > 0.05$).

3.1.2. *Cognitive flexibility*

Cognitive flexibility was significantly related to system control failures ($r_s = 0.34$, $p < 0.05$). According to this finding, participants with low cognitive flexibility tended to perform better on system control than did participants of high cognitive flexibility. An opposing pattern of correlations was found for diagnostic performance, with high cognitive flexibility being associated with accurate fault diagnosis ($r_s = -0.31$, $p < 0.05$) and fault identification time ($r_s = -0.35$, $p < 0.05$).

Examination of the subscales of the overall cognitive-flexibility scale (see Table 2) shows that the sub-scale 'constructing knowledge' in particular explains these correlations. Participants who took initiative and constructed their knowledge tended to perform better than those who rather recalled rigid prescriptions from memory in both accuracy of fault diagnosis ($r_s = -0.30$, $p < 0.05$) and fault identification time ($r_s = -0.34$, $p < 0.05$). The participants who took initiative and constructed their knowledge also seemed to gain a superior understanding of the system, although this correlation was only marginally significant ($r_s = 0.27$, $p = 0.052$). By contrast, system control was correlated to a cognitive style in which the individual prefers passively receiving information and knowledge to constructing knowledge ($r_s = 0.29$, $p < 0.05$).

3.1.3. *Self-efficacy and personality*

Contrary to the hypotheses, self-efficacy and the personality traits showed no significant relationships to process control performance. See Table 1 for corresponding correlation coefficients.

3.2. *Regression analysis with operator characteristics and performance*

A hierarchical regression analysis was conducted, with system control failures serving as the criterion and with cognitive ability, cognitive flexibility and system knowledge serving as predictors (Table 3). The selection of predictors and their order of entry was based on past research, as recommended by Field (2005). Cognitive ability was therefore entered first, as past work indicated that cognitive ability was a strong predictor of performance (Schmidt and Hunter 1998; see section 1.2.1). System knowledge was added as a further predictor because it also correlated with system control performance.

All predictors showed significant correlations with the criterion. Cognitive flexibility and system knowledge were significant predictors of system control performance ($p < 0.05$). Cognitive ability accounted for 8% of the variation in system control performance; cognitive flexibility, for 10%. Inclusion of system knowledge made it possible to explain an additional 11% of the variation in system control. Overall, the two operator characteristics, cognitive ability and cognitive flexibility, together with system knowledge account for almost 30% of the variation in system control performance.

Table 2. Spearman correlations between sub-scales of cognitive flexibility and performance.

Performance measure	Leaving complexity as it is	Constructing knowledge	Self-directed learning
Primary task performance			
System control failures	0.09	0.29*	0.14
Diagnostic errors	−0.09	−0.30*	−0.15
Fault identification time	−0.25	−0.34*	−0.10
System knowledge	0.05	0.27	−0.06

*$p < 0.05$.

The same hierarchical regression analysis was carried out with diagnostic performance as a criterion and with cognitive ability, cognitive flexibility and system knowledge as predictors. The analyses revealed no significant predictors for diagnostic errors, fault identification time (see Table 4), or the secondary task performance measures.

4. Discussion

The primary goal of this study was to improve understanding of the relationship between operator characteristics and process control performance. Cognitive ability and cognitive flexibility emerged as the best predictors of process control performance. An important finding was that there were differences between the predictors of system control performance and diagnostic performance. All these findings are relevant to operator selection.

There are two major findings in this study, which both have implications for personnel selection. The first major finding was that cognitive ability was a good predictor for process control performance and for diagnostic performance (though the association was only observed for novel faults and not for practised ones). Cognitive ability was also significantly correlated with system knowledge. With regard to system knowledge, Schmidt and Hunter (1998) argue that cognitive ability indirectly affects performance through the acquisition of job knowledge. Presumably, cognitive ability also indirectly affects system control performance by the same route, but the current sample was too small to permit an accurate examination of this possibility.

Table 3. Summary of hierarchical regression analysis with system control failures as a criterion.

Predictors of system control failure	B	$SE\ B$	β	Significance
Step 1				
Cognitive ability	−0.31	0.18	−0.28	0.093
Step 2				
Cognitive ability	−0.39	0.18	−0.36	0.034
Cognitive flexibility	1.76	0.87	0.33	0.051
Step 3				
Cognitive ability	−0.19	0.19	−0.17	0.327
Cognitive flexibility	1.70	0.83	0.32	0.047
System knowledge	−0.65	0.30	−0.37	0.036

$R^2 = 0.08$ for step 1; $\Delta R^2 = 0.10$ for step 2; $\Delta R^2 = 0.11$ for step 3.

Table 4. Summary of hierarchical regression analysis with fault identification time as a criterion.

Predictors of fault identification time	B	$SE\ B$	β	Significance
Step 1				
Cognitive ability	−0.89	0.75	−0.20	0.244
Step 2				
Cognitive ability	−0.65	0.76	−0.15	0.394
Cognitive flexibility	−5.10	3.77	−0.23	0.185
Step 3				
Cognitive ability	−0.40	0.87	−0.09	0.647
Cognitive flexibility	−5.18	3.81	−0.23	0.184
System knowledge	−0.82	1.37	−0.11	0.556

$R^2 = 0.04$ for step 1; $\Delta R^2 = 0.005$ for step 2; $\Delta R^2 = 0.01$ for step 3.

Cognitive ability and cognitive flexibility together with declarative knowledge account for nearly 30% of the variance in system control. The effect sizes in this study are similar to findings in the meta-analysis by Schmidt and Hunter (1998), who reported relationships of $r = 0.51$ between cognitive ability and work performance ($R^2 = 26\%$). The similar effect sizes also suggest some ecological validity of the present experiment since the present findings were comparable to results obtained in real work environments.

The second major finding of the study was that cognitive flexibility was a good predictor of process control performance, but the effects emerged in different directions for the two tasks fault diagnosis and system control. Operators with high cognitive flexibility outperformed operators with low cognitive flexibility on fault diagnosis, whereas the reverse was true for system control (i.e. low cognitive flexibility was associated with better performance). This suggests that the most appropriate level of cognitive flexibility is task-dependent to the extent that the positive effects of a certain level may not only be neutralised under a different task but might even turn to a disadvantage. This supports Landeweerd's (1979) finding that operator skills in system control and in fault diagnosis are independent of each other. Taken together, these findings show that performance of the two main tasks of process control are supported by different levels of cognitive flexibility. While these differences should be considered in personnel selection, they entail a dilemma since operators will be required to be good at both system control and fault diagnosis. There are several possibilities to deal with that dilemma. First, operators with a medium score in cognitive flexibility may be selected to avoid very poor performance on either of the two process control tasks. This approach favours generalists rather than specialists. However, while this would prevent the selection of operators who perform very poorly at one of the tasks, the downside would be that these operators will not excel at any of the tasks. Second, a task analysis is conducted to determine if any of the two tasks is more important than the other for overall performance in a specific work context. For instance, in some organisations operators are only expected to carry out system control tasks while senior colleagues are responsible for diagnostic tasks (Williams and Taylor 1993). In that case, levels of required cognitive flexibility could be adjusted accordingly. Third, if process control tasks are carried out in teams, a possible solution may be to build diverse teams whose members are matched for their cognitive style (e.g. high- and low-flexibility operators are members of the same team). This approach may be promising because cognitive diversity in teams has been shown to benefit system control performance and diagnostic performance (Sauer *et al.* 2006). Overall, the task dependency of cognitive flexibility adds an additional layer of complexity to the literature on cognitive styles, which has generally assumed that although 'bipolar dimensions represented two equally efficient ways of solving a task, in reality, one strategy was usually more effective than the other' (Kozhevnikov 2007, p. 466). Furthermore, the cognitive flexibility scale was especially developed for complex and ill-structured tasks, which has been identified as a particular need in the human–computer domain (Hockey 1990).

In addition to the demonstrated utility of the cognitive ability and cognitive flexibility tests, it is of special practical interest to note that these instruments are relatively inexpensive, quick and easy to administer. For example, the Wonderlic ability test takes 12 min to complete and the cognitive flexibility questionnaire about 5 min. Due to these advantages, the cognitive ability and cognitive flexibility tests compare favourably to other methods employed in personnel selection, such as interviews.

When looking at the overall pattern of association between operator characteristics and performance, some of the correlations predicted by the literature were not found in the present study. This may be due to environmental variables (e.g. presence of stressors) or the training approach. For example, one may presume that emotional stability as a personality factor might become more relevant with increasing work pressure and stress levels. Similarly, in a training approach as used in the present study, few degrees of freedom are provided to operators compared to other training approaches. This may generally lead to individual characteristics becoming less relevant (cf. Mischel 1968).

The influence of different interventions on the relationship between operator characteristics and performance needs to be examined to advance the understanding of the individual correlates of process control performance. With appropriately designed interventions, operators could be better supported, matching cognitive styles and abilities. For example, one may consider a training method in which the emphasis and training time on specific tasks is varied according to individual operator needs (cf. emphasis shift training; Gopher 2007). This may be in the form of exercises emphasising fault diagnosis (for operators with low cognitive flexibility because they are generally less good at this task) or system control (for operators with high cognitive flexibility because they are generally less good at this task).

Finally, recommendations for future research on operator characteristics are given. The present study was of an exploratory nature since empirical results on operator characteristics, especially cognitive flexibility, are quite scarce. Therefore, further research is needed to test whether the results found in this study can be replicated and whether they can be applied to different settings. For instance, the experiment involved trainee operators at the beginning of their vocational careers. This needs to be complemented with work using real operators to examine whether the results would hold true for more experienced operators.

Acknowledgements

This project has been partly funded by the Swiss National Science Foundation (No. P001–106354).

References

Barrick, M.R. and Mount, M.K., 1991. The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.

Bartram, D., 1995. The predictive validity of the EPI and 16PF for military flying training. *Journal of Occupational and Organizational Psychology*, 68, 219–236.

Brandimonte, M., Einstein, G.O., and McDaniel, M.A., 1996. *Prospective Memory: Theory and Applications*. Hillsdale: Erlbaum.

Colquitt, J.A., LePine, J.A., and Noe, R.A., 2000. Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678–707.

Debowski, S., Wood, R.E., and Bandura, A., 2001. Impact of guided exploration and enactive exploration on self-regulatory mechanisms and information acquisition through electronic search. *Journal of Applied Psychology*, 86, 1129–1141.

Field, A., 2005. *Discovering statistics using SPSS*, 2nd ed. London: SAGE Publications.

Gopher, D., 2007. Emphasis change as a training protocol for high-demand tasks. *In*: A.F. Kramer, D.A. Wiegmann, and A. Kirlik, eds. *Attention: From theory to practice*. Oxford: Oxford University Press, 209–224.

Hockey, G.R.J., 1990. Styles, skills and strategies: Cognitive variability and its implications for the role of mental models in HCI. *In*: D. Ackermann and M.J. Tauber, eds. *Mental models and human-computer interaction 1*. North-Holland: Elsevier, 113–129.

Hockey, G.R.J., Sauer, J., and Wastell, D.G., 2007. Adaptability of training in simulated process control: Comparison of knowledge- and rule-based guidance under task changes and environmental stress. *Human Factors*, 49, 158–174.

Hollnagel, E., 1998. *Cognitive reliability and error analysis method*. Oxford: Elsevier Science Ltd.
Hunter, D.R. and Burke, E.F., 1994. Predicting aircraft pilot training success: A meta-analysis of published research. *The International Journal of Aviation Psychology*, 4, 297–313.
International Atomic Energy Agency, 1996. *Nuclear power plant personnel training and its evaluation: A guidebook*. Executive summary. IAEA-TECDOC-1411. Vienna: International Atomic Energy Agency.
International Atomic Energy Agency, 2002. *Recruitment, qualification and training of personnel for nuclear power plants*. Safety Standard Series. No. NS-G-2.8. Vienna: International Atomic Energy Agency.
International Atomic Energy Agency, 2006. *Competency assessments for nuclear industry personnel*. Vienna: International Atomic Energy Agency.
Kluge, A., 2008. What you train is what you get? Task requirements and training methods in complex problem-solving. *Computers in Human Behavior*, 24, 284–308.
Kluge, A., *et al.*, in press. Designing training for process control simulators: A review of empirical findings and current practices. *Theoretical Issues in Ergonomics Science*.
Landeweerd, J.A., 1979. Internal representation of a process, fault diagnosis and fault correction. *Ergonomics*, 22, 1343–1351.
Mischel, W., 1968. *Personality and assessment*. New York: Wiley.
Morris, N.M. and Rouse, W.B., 1985. Review and evaluation of empirical research in troubleshooting. *Human Factors*, 27, 503–530.
Saucier, G., 1994. Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment*, 63, 506–516.
Sauer, J., Hockey, G.R.J., and Wastell, D.G., 2000a. Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics*, 43, 2043–2064.
Sauer, J., Wastell, D.G., and Hockey, G.R.J., 2000b. A conceptual framework for designing microworlds for complex work domains: A case study of the Cabin Air Management System. *Computers and Human Behavior*, 16, 45–58.
Sauer, J., *et al.*, 2006. Cognitive diversity and team performance in a complex multiple-task environment. *Ergonomics*, 10, 934–954.
Sauer, J., *et al.*, 2008. The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics*, 51, 953–967.
Schmidt, F.L. and Hunter, J.E., 1998. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
Schyns, B. and Collani, G., 2002. A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology*, 11, 219–241.
Spiro, R.J., Feltovich, P.J., and Coulson, R.L., 1996. Two epistemic world-view: prefigurative schemas and learning in complex domains. *Applied Cognitive Psychology*, 10, 51–61.
Stajkovic, A.D. and Luthans, F., 1998. Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124, 240–261.
Stanton, N. and Ashleigh, M., 1996. Selecting personnel in the nuclear power industry. *In*: N. Stanton, ed. *Human factors in nuclear safety*. London: Taylor & Francis, 159–186.
Wickens, C.D., Gordon, S.E., and Liu, Y., 1998. *An introduction to human factors engineering*. New York: Addison Wesley.
Wickens, C.D. and Hollands, J.G., 2000. *Engineering psychology and human performance*, 3rd ed. Upper Saddle River: Prentice Hall.
Williams, J.C. and Taylor, S.E., 1993. *A review of the state of knowledge and of current practice in selection techniques for process operators*, HSE contract research report No. 58. London: HMSO.
Wittmann, W.W. and Hattrup, K., 2004. The relationship between performance in dynamic systems and intelligence. *Systems research and Behavioural Science*, 21, 393–409.
Wonderlic, 2002. *Wonderlic Personnel Test*. Libertyville: Wonderlic Inc.

Chapter 5

Study II: Cognitive and personality variables of operators

Burkolter, D., Kluge, A. & Brand, M. (2009). Cognitive and personality variables of operators: individual characteristics and process control performance. In D. de Waard, J. Godthelp, F.L. & K.A. Brookhuis (Eds.). *Human factors, security and safety* (pp. 77–88). Maastricht: Shaker Publishing.

Reprinted with permission by Shaker Publishing BV.

Cognitive and personality variables of operators: individual characteristics and process control performance

Dina Burkolter, Annette Kluge, & Matthias Brand
University of Duisburg-Essen
Duisburg, Germany

Abstract

Two studies analysing the relationships between operator characteristics and process control performance were conducted in order to replicate and extend earlier research by Burkolter et al. (2007, 2009). The first study (N = 41) examined the link between general mental ability (GMA), cognitive flexibility and conscientiousness on the one hand, and performance in a simulated process control task on the other. GMA correlated significantly with diagnostic performance and knowledge, while cognitive flexibility and conscientiousness were not related to performance. The second study (N = 50) incorporated working memory capacity, set-shifting and decision-making as well as the personality variables Need for Cognition (NFC) and perfectionism. Working memory capacity and set-shifting were related to diagnostic performance, while decision-making (risky decisions) was associated with both system control and diagnostic performance. Moreover, NFC and declarative knowledge were significantly correlated. Confirming earlier results, GMA exhibited relations to diagnostic performance and declarative knowledge. Practical implications as well as implications for future research are discussed.

Introduction

The present studies aim to continue and extend previous research results, which addressed relationships between individual characteristics and performance in early stages of skill acquisition (see Burkolter et al., 2007, 2009; Kluge et al., submitted).

Altogether, these studies form part of a comprehensive research programme that links cognitive requirements of process control tasks to training. When designing training, it is necessary to analyse the cognitive requirements of a task (e.g. through cognitive task analysis) in order to derive training objectives and criteria (Burkolter et al., 2007). However, the amount of learning and the training results might be limited by individual characteristics. Therefore, the overriding research question focuses on which aspects of process control performance are affected by training, and how training can best support learning and performance with respect to individual characteristics. Our underlying assumption is that process control performance is composed of a dynamic interplay between cognitive processes and

In D. de Waard, J. Godthelp, F.L. Kooi, and K.A. Brookhuis (Eds.) (2009). *Human Factors, Security and Safety* (pp. 77 - 88). Maastricht, the Netherlands: Shaker Publishing.

individual characteristics. Therefore, before developing training scenarios and methods, it is necessary to investigate the elements of a process control task that are affected by person-related variables and address the question of which of these are trainable.

Our earlier research showed that general mental ability (GMA) was positively related to system control performance and declarative knowledge (knowledge about the system and the relationships between parameters) as well as to diagnostic performance of novel faults, i.e. not previously encountered system faults. These findings confirmed the results of earlier studies (e.g. Schmidt & Hunter, 1998; Morris & Rouse, 1985). Furthermore, cognitive flexibility was found to be a good predictor of process control performance. Cognitive flexibility as a construct distinguishes between a reductive and an expansive cognitive style. Individuals with high cognitive flexibility have a flexible and situation-adaptive assembly of knowledge, while those with low cognitive flexibility prefer simplicity and rigid prescriptions from memory (Spiro et al., 1996). Interestingly, the effect of cognitive flexibility was found to occur in different directions. Operators with high cognitive flexibility performed best on diagnostic performance, whereas operators with low cognitive flexibility performed best on system control. GMA, cognitive flexibility and declarative knowledge accounted for about 30% of the variability of system control. A second study (Kluge et al., submitted) with a similar sample showed comparable results, and corroborated the impact of cognitive flexibility on process control performance. Finally, regarding personality traits, no correlation was found between conscientiousness and process control performance (Burkolter et al., 2009), despite the fact that a meta-analysis by Barrick and Mount (1991) had suggested such a relationship.

Although these findings demonstrated the relevance of operator characteristics in terms of explaining variance in performance, they also uncovered the need to (possibly) confirm previous results and address individual differences more specifically. We therefore sought to extend this type of research by investigating the interplay between GMA, cognitive style and task-related personality traits in different samples, for instance with regard to age and experience. These differing samples were selected based on the insight that in order to support operators most effectively, training needs to be tailored according to the trainees' education, experience and abilities.

Moreover, in addition to the basic constructs such as GMA, cognitive style and personality traits, there is a need to examine more specific cognitive variables and personality traits, which might explain a higher amount of variance in complex task performance than the basic concepts. Based on empirical findings summarised below, this study began by addressing Need for Cognition (NFC), perfectionism, working memory capacity, and set-shifting performance as a promising set of individual characteristics that are expected to have an impact on process control performance: NFC refers to "the tendency for an individual to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). In a recent study by Day et al. (2007), NFC was shown to be related to performance of a complex skill. Facets of perfectionism such as concern over mistakes and personal standards were found to be related to decision-making (Brand & Altstötter-Gleich, 2008), and working memory capacity was associated with performance in complex systems characterised by a high amount of interconnectivity (e.g. Wittmann & Hattrup, 2004). Set-shifting (or categorisation) as a sub-component of executive functions, is assumed to be related to process control performance, as it has been found to be relevant for problem-solving (Hockey, 1990). Finally, decision-making under uncertainty was included in the study as an important factor in dealing with complex systems (e.g. Wickens et al., 1998).

The setting

Two studies were conducted in order to investigate the relationships between cognitive and personality variables and process control performance. Both studies involved a simulated process control environment called Cabin Air Management System (CAMS), which has already been employed in a number of previous studies (for details, see Sauer et al., 2000). By employing a simulation, we chose a compromise between experimental control and realism, under the assumption that the simulation represents the essential characteristics of a real-world environment (Gonzalez et al., 2005; Sauer et al., 2008). CAMS simulates the automated life-support system of a spacecraft, but its underlying principles correspond to a process control task. Accordingly, the two main tasks of the operator are to maintain five parameters in a predefined zone and to intervene in the case of a system fault. System control performance was measured as the duration with which one or more of the parameters lay outside of the predefined zone, converted into percentages. Diagnostic performance included the two measures diagnostic errors (percentage of incorrect fault diagnoses) and diagnosis time (required time for correct fault diagnosis in seconds). Knowledge was assessed using a written questionnaire that contained items relating to the parameters, processes and relationships in CAMS (declarative knowledge) as well as items regarding fault descriptions, fault symptoms and repair steps (procedural knowledge). The knowledge test was composed of multiple-choice items as well as open questions (for details see Sauer et al., 2008; Kluge et al., submitted).

The procedure was similar in both studies. Cognitive and personality variables were assessed before the training session either through questionnaires or using computer-based tests. Participants in both studies were trained in the use of CAMS for several hours, including an introduction by means of multimedia-based instructions and a main training session on the control and diagnosis of five system faults. Training was carried out in groups of typically five to ten participants, supervised by one instructor.

Study 1

The goal of the first study was to replicate earlier findings and examine the relationship between operator characteristics and performance more deeply in different samples, i.e. by inviting participants of different ages and with different levels of education.

Participants and methods

Forty-one engineering students (four female) from universities of applied sciences in Switzerland took part in the study. The mean age of the participants was 24.7 years (SD = 4.0), with ages ranging from 19 to 35 years. They were paid approximately € 60 (100 CHF) for participation.

At the beginning of the training session, which lasted five hours in total, different questionnaires were administered (approx. 30 minutes). Cognitive flexibility was measured using the 'Cognitive Flexibility Inventory' (Spiro et al., 1996), GMA using the 'Wonderlic Personnel Test' (Wonderlic Inc, 2002), and items measuring conscientiousness were extracted from Saucier's (1994) 'Big Five Markers'. Following this, participants were introduced to the CAMS task through multimedia-based instructions, and were provided with training and time to practise the process control task. Results are reported for the test (70 min) that was conducted two weeks after the first session. Knowledge was assessed using a paper-and-pencil questionnaire.

Results

Descriptive statistics on all performance variables as well as personality variables are provided in Table 1. While there were significant relationships between GMA and performance, cognitive flexibility and conscientiousness were not significantly associated with process control (see Table 2).

Table 1. Descriptive statistics on performance and personality variables

Variables	*M*	*SD*
Process control performance		
System control failures (%)	27.05	9.36
Diagnostic errors (%)	53.66	21.89
Diagnostic speed (s)	330.07	66.91
Knowledge (max. 14 pts)	6.76	1.97
Personality variables		
GMA[1] (max. 50 pts)	25.61	5.65
Cognitive flexibility (–4 to +4)	0.77	1.07
Conscientiousness (1-9)	6.71	1.09

Notes: N = 41, [1]General mental ability

Table 2. Correlations between personality and performance variables

Performance measures	*GMA[1]*	*Cognitive flexibility*	*Conscientiousness*
Process control performance			
System control failures	–.28	–.27	–.31
Diagnostic errors	$-.47^{**}$	–.16	–.21
Diagnostic speed	$-.43^{**}$	–.12	–.11
Knowledge	$.33^{*}$	.15	.03

Notes: N = 41; $^{*}p < .05$, $^{**}p < .01$; [1]General mental ability

GMA was significantly correlated with both measures of diagnostic performance (diagnostic errors and diagnostic speed, see Table 2) as well as acquired knowledge.

The negative correlations indicate that as a tendency, the higher the participants' GMA scores, the better their performance. Neither cognitive flexibility nor conscientiousness was significantly related to process control performance.

Summary

In part, the present results confirmed earlier findings discussed in the Introduction. In line with previous research, GMA was significantly correlated with diagnostic performance and knowledge. Similarly to other work environments (Schmidt & Hunter, 1998), GMA also supports performance with complex systems. As in the previous study, conscientiousness was not significantly related to process control performance. However, there were also differences compared to the previously obtained results, for instance regarding cognitive flexibility. While previous research showed a relationship between cognitive flexibility and process control performance, these findings were not confirmed in the present study.

Study 2

The purpose of the second study was to include additional variables in order to assess their incremental predictive qualities in terms of process control performance. The supplementary variables were derived from studies from neuropsychology and cognitive science, especially research concerning executive functions. Working memory capacity, set-shifting performance and decision-making, assessed with computer-based tests, were included as cognitive variables. As personality variables, NFC and perfectionism were additionally incorporated. The cognitive variables as well as the personality variables were shown in earlier research to be related to performance in complex systems.

Participants and methods

Fifty engineering students (26 female) from a university in Germany took part in the study. Their mean age was 23.6 years (SD = 2.9), with ages ranging from 20 to 34 years. Participants were paid 100 EUR to attend the experiment, which was held over three sessions.

At the first session, person-related variables were assessed (approx. one hour). Working memory capacity was assessed using the 'n-back' task with one-digit numbers, presented in a random sequence. In this n-back task, the participant monitors a series of numbers and is required to indicate whenever a number is presented that was presented two or three trials previously (see Owen et al., 2005). Ten stimulus blocks (two practice blocks and eight experimental blocks) with 24 stimulus trials each were given, alternating between 2-back and 3-back conditions (only 2-back data were analysed here; see Schoofs et al., 2008). Set-shifting performance was measured by the 'Modified Card Sorting Test' (Nelson, 1976), in which the participant is required to sort a deck of cards according to a particular rule. Feedback regarding whether the correct rule has been applied is given after each card has been sorted. To assess the participants' decision-making, the 'Game of Dice Task' (Brand et al., 2005) was employed. In this game, a fictitious starting capital has to be increased within 18 throws of a die by choosing numbers or a combination of numbers that are related to different winning probabilities (i.e. gains or losses). All of the assessments mentioned were carried out using computer-based tasks. Furthermore, different questionnaires on personality were administered: Perfectionism was measured with the German version of the 'Multidimensional Perfectionism Scale' (Altstötter-Gleich & Bergemann, 2006), 'Need for Cognition' was examined using the German version of the 'Rational-Experiential Inventory' (Keller et al., 2000), cognitive flexibility with the 'Cognitive Flexibility Inventory' (Spiro et al. 1996), GMA with the 'Wonderlic Personnel Test' (Wonderlic Inc, 2002), and conscientiousness by Saucier's (1994) Big Five Markers.

At the second session, participants were introduced to the experimental task and provided with training and some time to practise (approx. four hours). The training was followed by a first immediate testing session and a second testing session one week later (both 70 min). Knowledge was assessed at the second testing session (Sauer et al., 2008; Kluge et al., submitted). Results are reported from the second testing session one week later.

Results

Compared to Study 1, participants in Study 2 scored similarly on GMA but differed with regard to cognitive flexibility (see Table 3). They also differed in terms of their absolute process control task performance. The overall pattern of results revealed significant relationships between GMA as well as NFC and process control performance (see Table 4) as well as significant correlations between working memory capacity, set-shifting performance, decision-making and process control (see Table 5).

Table 3. Descriptive statistics on performance, personality and cognitive variables

Variables	M	SD
Process control performance		
System control failures (%)	44.78	15.76
Diagnostic errors (%)	45.33	25.65
Diagnostic speed (sec)	333.31	63.37
Knowledge (max. 33 pts)	15.75	5.59
Personality variables		
GMA (max. 50 pts)	25.78	6.28
Cognitive flexibility (–4 to +4)	–2.39	0.38
Conscientiousness (1-9)	5.95	1.30
Need for cognition (1-7)	5.27	0.55
Perfectionism[2] (1-6)	1.90	0.95
Cognitive variables		
Working memory: Correct (%)	81.77	15.30
Set-shifting: Correct (max. 48)	40.50	7.11
Decision-making: Risky decisions (max. 18)	1.20	1.87

Notes: N = 50, [1]General mental ability; [2]subfacet: doubts about actions

GMA was significantly correlated with diagnostic performance (diagnostic errors, see Table 4) and system knowledge. Working memory capacity was significantly related to both measures of diagnostic performance. This negative correlation indicates as a tendency that participants scoring high on the working memory test committed fewer diagnostic errors and needed less time to diagnose a fault state correctly.

Table 4. Correlations between personality variables and process control performance

Performance measures	*GMA*[1]	*Cognitive flexibility*	*Conscien-tiousness*	*Need for cognition*	*Perfect-ionism*[2]
Process control performance					
System control failures	–.20	.13	–.11	–.14	–.01
Diagnostic errors	–.28*	–.16	–.08	–.21	.19
Diagnostic speed	–.14	–.18	–.15	–.21	.20
Knowledge	.43**	.12	.15	.40**	–.27

Notes: N = 50; $^{*}p < .05$, $^{**}p < .01$; [1]General mental ability; [2]subfacet: doubts about actions

Cognitive flexibility was not significantly related to process control performance, but the tendency was similar to results obtained in earlier research (see Introduction). Regarding NFC, a significant relationship with system knowledge was found, but not with the other performance measures. Again, conscientiousness was not associated with the performance criteria. Finally, perfectionism was not significantly related to process control performance, except for a marginally significant relationship between knowledge and the perfectionism facet 'doubts about actions' ($r = -.27$, $p = .058$).

Table 5. Correlations between cognitive variables and process control performance

Performance measures	*Working memory: Correct*	*Set-shifting: Correct*	*Decision-making: Risky decisions*
Process control performance			
System control failures	–.14	–.08	.32*
Diagnostic errors	–.34*	–.36*	.30*
Diagnostic speed	–.30*	–.23	.30*
Knowledge	.16	.40**	–.23

Notes: N = 50; $^{*}p < .05$, $^{**}p < .01$

Regarding set-shifting performance, there was a significant relationship with diagnostic performance in terms of diagnostic errors as well as knowledge. Significant correlations were observed between decision-making and both system control and diagnostic performance (diagnostic errors and diagnostic speed, see Table 5). Thus, operators making more risky decisions tended to perform lower in terms of process control.

Hierarchical regression analyses were carried out in order to reveal whether, and how much more, variance in performance could be explained by several variables together compared to a single variable alone. The regression analyses were conducted with the variables that correlated most highly with the performance measures chosen as predictors. Thus, two regression analyses were conducted with diagnostic errors and diagnostic speed as criteria and working memory capacity as well as decision-making as predictors (see Table 6). Working memory and decision-making were both significant predictors of diagnostic performance. Working memory accounted for 11% of variation in diagnostic errors and 9% of variation in diagnostic speed. Overall, 22% of the variation in diagnostic errors and 20% of variation in diagnostic speed was explained by the two predictors working memory capacity and decision-making.

Table 6. Summary of hierarchical regression analyses with diagnostic errors and diagnostic speed as criteria

Predictors of diagnostic errors	*B*	*SE B*	*β*	*Significance*
Step 1				
Working memory	–0.56	0.23	–.34	.019
				Step 2
Working memory	–0.62	0.22	–.38	.007
Decision-making: Risky decisions	4.52	1.76	.34	.014
		$R^2 = .11$ for step 1. $\Delta R^2 = .11$ for step 2		
Predictors of diagnostic speed				
Step 1				
Working memory	–1.23	0.59	–.30	.041
Step 2				
Working memory	–1.40	0.56	–.34	.017
Decision-making: Risky decisions	11.09	4.54	.33	.018
		$R^2 = .09$ for step 1. $\Delta R^2 = .11$ for step 2		

A third hierarchical regression analysis was performed with acquired knowledge as a criterion and GMA and NFC as predictors (see Table 7). GMA accounted for 19% of the variation in knowledge alone, but including NFC made it possible to explain a further 8%, so that in total, 27% of the variation was accounted for by these two predictors alone.

Table 7. Summary of hierarchical regression analysis with knowledge as a criterion

Predictors of knowledge	*B*	*SE B*	*β*	*Significance*
Step 1				
GMA	0.39	0.12	.43	.002
Step 2				
GMA	0.30	0.12	.34	.015
Need for cognition	3.02	1.36	.30	.032

Notes: $R^2 = .19$ for step 1. $\Delta R^2 = .08$ for step 2

Summary

There were several results that go beyond earlier findings, such as the positive relationship between NFC and knowledge. Furthermore, GMA and NFC were significant predictors of knowledge, and accounted for 27% of variance in knowledge acquisition. Confirming earlier results, Study 2 showed that GMA was related to the correct diagnosis of system faults as well as the amount of acquired knowledge.

As for the cognitive variables, working memory capacity and decision-making (i.e. risky decisions) were related to diagnostic performance, and moreover, they were significant predictors of diagnostic performance. 20% of the variation in diagnostic performance could be explained by working memory and decision-making. In view of all possible influences on performance (e.g. experience or work design), the amount of variance explained by just a few person-related variables is considerable.

As in Study 1, there were no significant relationships either between cognitive flexibility and the performance measures or between conscientiousness and performance, even though conscientiousness has been shown to be a strong predictor of job performance for a range of occupations (Barrick & Mount, 1991) in general. However, this does not seem to be the case in terms of process control.

Discussion

The goal of the conducted studies was to continue and extend research on the relationships between individual characteristics and process control performance in order to gain more insights for personnel selection and training design. One result emerged as the most stable across the different samples: GMA has consistently been shown to be related to process control performance, especially knowledge, independently of the sample. This finding is in line with research by Schmidt and Hunter (1998) or Morris and Rouse (1985).

The inclusion of executive functions such as working memory capacity and set-shifting performance revealed additional interesting results. Working memory capacity and set-shifting performance were found to be of relevance for diagnostic performance and knowledge acquisition, but not for system control performance. Future research should investigate the relationships between executive functions and process control further by integrating other executive functions such as goal-setting, selection and implementation of strategies or control of actions (see Hockey, 1990). Various selection criteria such as problem-solving abilities or communication skills are suggested by the IAEA (2002), but the assessment of executive functions such as set-shifting performance or working memory capacity is not among the recommendations. It might be of interest for personnel selection in process industries that the computer-based tools employed to assess working memory capacity and set-shifting performance are relatively easy and quick to administer. Furthermore, knowledge on individual characteristics might be employed to tailor training appropriately to trainees, for example with CBT, which can be matched to different learning and cognitive styles by presenting information accordingly (cf. Russell, 1997).

The study findings give rise to the suggestion of examining the relationships between individual characteristics and performance closely regarding age or experience of individuals. Significant relationships between process control performance and cognitive flexibility were found for apprentices, but not for students (although apprentices and students were comparable regarding GMA, see also Burkolter et al., 2007). A possible explanation for this finding could be that the cognitive style may be relevant for apprentices to guide their learning and application process, while this might be less relevant for students, who may be more aware of the advantages and disadvantages of their cognitive styles and might have acquired compensation strategies. Although cognitive styles are seen as fixed attributes, individuals can nevertheless "develop strategies to maximise their strength and minimise weaknesses once aware of their own style" (Russell, 1997, p. 208). However, as age and education (apprentices vs. students) are confounded in this case, it cannot be determined whether age or education in particular might have influenced the observed difference.

To put the findings in an appropriate context, it has to be considered that the studies involved engineering students and a simulated process control task instead of experienced operators in a real work setting. This might limit the validity of the findings for process industries. Particularly in light of the fact that our first studies regarding operator characteristics and the employed task showed differences between samples, further research with more experienced operators from process industries is needed in order to investigate under what conditions findings also apply to more experienced workers. However, even experienced operators start out by learning the fundamentals and building up a generic mental model on how a plant works, upon which, for example, they later build their situational model (Vicente et al., 2004). The situational model then forms expectations regarding how the plant is going to act or react. Studying performance in the early stages of skill acquisition thus holds special importance for supporting operators.

References

Altstötter-Gleich, C. & Bergemann, N. (2006). Testgüte einer deutschsprachigen Version der Mehrdimensionalen Perfektionismus Skala von Frost, Marten, Lahart und Rosenblate (MPS-F). [Statistics and validity of a German version of the Multidimensional Perfectionism Scale of Frost, Marten, Lahart and Rosenblate (MPS-F)]. *Diagnostica, 52*, 105–118.

Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.

Brand, M. & Altstötter-Gleich, C. (2008). Personality and decision-making in laboratory gambling tasks – evidence for a relationship between deciding advantageously under risk conditions and perfectionism. *Personality and Individual Differences, 45*, 226–231.

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H.J. (2005). Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: associations with executive functions. *Neuropsychology, 19*, 267–277.

Burkolter, D., Kluge, A., Sauer, J., & Ritzmann, S. (2009). The predictive qualities of operator characteristics for process control performance: The influence of personality and cognitive variables. *Ergonomics, 52*, 302-311.

Burkolter, D., Kluge, A., Schüler, K., Sauer, J., & Ritzmann, S. (2007). Cognitive requirement analysis to derive training models for controlling complex systems. In D. de Waard, G.R.J. Hockey, P. Nickel, and K.A. Brookhuis (Eds.), *Human Factors Issues in Complex System Performance* (pp. 475–484). Maastricht: Shaker Publishing.

Cacioppo, J.T. & Petty, R.E. (1982). The need for cognition. *Personality and Social Psychology, 42*, 116–131.

Day, E.A., Espejoa, J., Kowollka, V., Boatmana, P.R., & McEntire, L.E. (2007). Modeling the links between need for cognition and the acquisition of a complex skill. *Personality and Individual Differences, 42*, 201–212.

Gonzalez, C., Vanyukov, P. & Martin, M.K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21*, 273–286.

Hockey, G.R.J. (1990). Styles, skills and strategies: Cognitive variability and its implications for the role of mental models in HCI. In D. Ackermann and M.J. Tauber (Eds.), *Mental models and human-computer interaction 1* (pp. 113–129). North-Holland: Elsevier.

Keller, J., Bohner, G. & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung - verschiedene Prozesse? Präsentation einer deutschen Fassung des «Rational-Experiential Inventory» sowie neuer Selbstberichtskalen zur Heuristiknutzung. [Intuitive and heuristic decision making – different proceses? Presentation of a German version of the 'Rational-Experiential Inventory' and new self report scales for heuristic use]. *Zeitschrift für Sozialpsychologie, 31*, 87–101.

Kluge, A., Ritzmann, S., Burkolter, D., & Sauer, J. (submitted). The interaction of drill and practice and error training with individual differences.

Morris, N.M. & Rouse, W.B. (1985). Review and evaluation of empirical research in troubleshooting. *Human Factors, 27*, 503–530.

Nelson, H. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*, 313–324.

Owen, A.M., McMillan, K.M., Laird, A.R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59.

Russell, A.J. (1997). The effect of learner variables cognitive style on learning performance in a vocational training environment. *Educational Psychology, 17*, 195–208.

Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment, 63*, 506–516.

Sauer, J., Burkolter, D., Kluge, A., Ritzmann, S., & Schüler, K. (2008). The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics, 51*, 953–967.

Sauer, J., Wastell, D.G., & Hockey, G.R.J. (2000). A conceptual framework for designing microworlds for complex work domains: A case study of the Cabin Air Management System. *Computers and Human Behavior, 16*, 45–58.

Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schoofs, D., Preuss, D. & Wolf, O.T. (2008). Psychosocial stress induces working memory impairments in an *n*-back paradigm. *Psychoneuroendocrinology, 33*, 643-653.

Spiro, R. J., Feltovich, P.J., Coulson, R.L. (1996). Two epistemic world-views: prefigurative schemas and learning in complex domains. *Applied Cognitive Psychology, 10*, 51–61.

Vicente, K.J., Mumaw, R.J., & Roth. E.M. (2004). Operator monitoring in a complex dynamic work environment: a qualitative cognitive model based on field observations. *Theoretical Issues in Ergonomic Science, 5*, 359–384.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An introduction to human factors engineering*. New York: Addison Wesley.

Wittmann, W.W. & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioural Science, 21*, 393–409.

Wonderlic (2002). *Wonderlic Personnel Test*. Libertyville: Wonderlic Inc.

Chapter 6

Study III: Comparative study of three training methods

Burkolter, D., Kluge, A., Sauer, J. & Ritzmann, S. (2010). Comparative study of three training methods for enhancing process control performance: emphasis shift training, situation awareness training, and drill and practice. *Computers in Human Behavior, 26*, 976–986.

Author's personal copy

Computers in Human Behavior 26 (2010) 976–986

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Comparative study of three training methods for enhancing process control performance: Emphasis shift training, situation awareness training, and drill and practice

Dina Burkolter [a,*], Annette Kluge [a], Jürgen Sauer [b], Sandrina Ritzmann [c]

[a] Business and Organizational Psychology, Department of Computational and Cognitive Sciences, Faculty of Engineering, University of Duisburg-Essen, Lotharstraße 65, D-47057 Duisburg, Germany
[b] Department of Psychology, University of Fribourg, Rue de Faucigny 2, CH-1700 Fribourg, Switzerland
[c] Research Institute for Organizational Psychology, University of St. Gallen, Varnbüelstrasse 19, CH-9000 St. Gallen, Switzerland

ARTICLE INFO

Article history:
Available online 9 March 2010

Keywords:
Training
Process control
Attention
Emphasis shift
Situation awareness
Practice

ABSTRACT

Three training methods to improve attention management skills in process control were compared. Forty students from technical disciplines participated in a five-hour module of emphasis shift training (EST), EST combined with situation awareness training (EST/SA), and drill and practice (D&P) on a simulated process control task. Participants were then tested three times for 45 min each (immediately after training, two weeks after training, and six weeks after training) for system control performance and diagnostic performance on familiar and nonfamiliar fault states. D&P led to superior diagnostic performance on familiar system faults. EST/SA training supported the diagnosis of novel system faults. EST was less effective than expected for system control performance. Implications for training design in process control are discussed.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Highly automated installations such as refineries or nuclear power plants involve extremely complex, dynamic process control tasks that require personnel to monitor and control the system and to detect, diagnose, and rectify malfunctions or make repairs (Kluge, Sauer, Schüler, & Burkolter, 2009). These tasks demand different kinds of attention from process control operators such as focused, divided, and selective attention (Wickens & McCarley, 2008). But as crucial as attention capacities are to process control performance, they are limited. The operator "would gain most if he or she could fully attend to all elements, at all times. However, such full attention is not possible. Hence, some priorities and tradeoffs must be established along with attention allocation strategies" (Gopher, 1996, p. 28). In other words, process control operators must learn to apportion their attention strategically. Attention is strongly associated with mental workload. Workload is on the one hand determined by exogenous task demands such as task difficulty and task priority, and on the other hand by endogenous supply of attentional or processing resources to support information processing (Tsang & Vidulich, 2006). Our research goal was to support novice learners by providing them with attention-management strategies in order to reduce their mental workload. Highly complex and demanding tasks such as process control can present demands which are difficult to cope with (Gopher, Weil, & Siegel, 1989). Therefore, especially in the early stages of training, learners should be supported by providing them with attention-management strategies.

1.1. Three training approaches to improve attention management and performance

Adapting, applying, and comparing auspicious training approaches drawn mostly from aviation, in this study we seek to improve attention, attention management, and, ultimately, performance on highly complex, dynamic process control tasks. In the following, we describe the approaches, their underlying theoretical concepts and why they are selected for learning a process control task. We introduce the training approaches of emphasis shift training (EST), situation awareness (SA) training, and drill and practice (D&P) (see Table 1) and summarize research findings on their effectiveness.

1.1.1. Emphasis shift training (EST)

The first training approach we selected for our research on attention management in process control was EST. It was originally developed by Gopher et al. (1989) to sharpen the ability to cope with highly demanding tasks and, especially, to strengthen attention management and the control of attentional resources. In EST,

* Corresponding author. Tel.: +49 203 3794207; fax: +49 203 3793411.
E-mail address: dina.burkolter@uni-due.de (D. Burkolter).

0747-5632/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2010.02.011

Author's personal copy

multiple changes in the emphasis (priority) on components of a task are introduced, but the whole task is left intact. Only the attention status of the subtasks is changed. Hence, EST is a part-task training approach. Key constructs of EST are strategies of performance, response schemas and the voluntary control of attention. Strategies and organized sets of response schemas are central to complex tasks. A strategy is a distinct approach of an individual to cope with the set of subgoals of a task. Strategies are controlled at the beginning, but may become high-level schemas that can be triggered automatically with training and practice. Once a schema is developed, the operation of it is assumed to require few attentional resources. Hence, attentional resources are freed for other tasks (Gopher et al., 1989). This is important, as operators performing complex tasks are required to coordinate many complex action sequences and subtasks in parallel. To support the development of strategies and schemas, Gopher et al. introduce the idea of voluntary control of attentional resources. There is theoretical and empirical evidence that attention control and attention management can be treated as a skill, and thus can be improved by training (Gopher, Weil, & Bareket, 1994). Gopher et al. (1989), for example, showed that spontaneous strategies developed by learners to try to cope with complexity were not very successful. In contrast, in EST, in which learners were provided with strategies, participants showed higher performance. Hence, trainees can be provided with strategies both to reduce mental workload and to improve performance.

EST is assumed to prepare participants for another challenge in process control. Some process control tasks, such as shutdown, start-up, and fault-finding, require completion only at certain intervals (Sauer, Hockey, & Wastell, 2000). Skill components are called upon not only in practiced, familiar situations but in novel, unfamiliar ones as well (Kluge et al., 2009). Training of unexpected, novel fault states should focus on attention-management strategies, because these strategies are central to responding to novel situations (Shebilske, Goettl, & Garland, 2000).

EST has thus far been used in different contexts, such as complex and dynamic environments (Space Fortress game), flying with a helmet-mounted display, touch-typing skills, and basketball (Gopher, 2007), but not to a process control task. EST makes it possible to resolve difficulties known from traditional part-task training approaches (Gopher et al., 1989). EST has been effective overall, especially for strengthening attention-management strategies (Gopher et al., 1994; Shebilske et al., 2000). EST has improved transfer of skills to new and changed tasks. However, EST's effectiveness has usually been tested at the end of training (Gopher, 2007), not after an extended retention interval. Therefore, there is a need to test whether EST can support skill retention.

Gopher et al. (1989) explain EST's effectiveness in terms of load reduction that permits a person to increase the resources invested in the learning of other tasks. They maintain that EST helps participants broaden their perspective of their given task, expand their knowledge about the efficiency of their own resources, and gain flexibility in adopting different modes of response that suit their individual capabilities.

1.1.2. SA training

Attention is also critical to achieving SA (Endsley, 1995b; Tsang & Vidulich, 2006). SA is understood to be the perception and comprehension of elements in the environment and the projection of their status in the near future. Research on SA goes back to aviation (e.g. Endsley, 1995b), but in recent years other fields have followed, including process control (e.g. Hogg, Follesø, Strand-Volden, & Torralba, 1995). In process control, operators have to monitor plant states, alarm screens and panels, and to observe the state of numerous system parameters and patterns among them in order to gather information about the functioning of the system and future process state changes (Endsley, 1995b; Vicente, Mumaw, & Roth, 2004). Limited attention capacities, such as lapses in attention and the constraints on the ability to accurately perceive several items in parallel, present a major limit to SA (Endsley, 1995b; Wickens & McCarley, 2008). Schemas can support individuals to develop SA in that they are mechanisms for directing attention in the perception process. SA is achieved by recognizing critical cues in the environment that will map to key characteristics of the schema or mental model. Schemas and mental models are developed through training and experience by noticing recurrent situational components and causal relationships (Endsley, 1995b). As SA is critical to process control performance (Endsley, 1995b; Wickens & McCarley, 2008), we selected SA training as a second training approach. To improve individuals' SA, training of attention sharing and task management strategies is recommended (Endsley & Robertson, 2000). By combining EST with SA training, an approach to enhance SA and performance both directly (with the SA training) and indirectly by training attention and task management to support SA (with the EST) was chosen.

Relatively few programs include the evaluation of SA training (Endsley & Robertson, 2000). However, Saus et al. (2006) have found empirical evidence substantiating the effects of the SA training

Table 1
Comparison of emphasis shift training, situation awareness training, and drill and practice.

Description	Emphasis shift training (Gopher, 2007; Gopher et al., 1989)	Situation awareness training (Endsley and Robertson 2000; Saus et al., 2006)	Drill and practice (Carlson et al., 1989; Ericsson et al., 1993)
Rationale	Learning through priority changes on subcomponents of a whole task	Learning through randomized "freezing" of a task with situation awareness questions and debriefing	Learning through rehearsal, repetition and practice of a task
Original application area	Aviation, complex tasks	Aviation, police, complex tasks	Nondynamic tasks, motor tasks
Empirical findings	Enhancement of attention-management strategies, useful for transfer to novel situations	Useful for individual situation awareness as well as performance, but few empirical studies on effects of situation awareness training	Useful for procedural tasks and longer retention intervals for familiar situations
Explanations for training effectiveness	Reduction of load allowing to invest more resources in learning other tasks	Improved competence to make decisions and project events in the future	Reduction of load on working memory
	Broader perspective of task, better knowledge of own resources	Enhanced mental models allowing for better understanding of situations	Increased speed of component processes and restructuring in the use of working memory
Usefulness assumed for	System control performance	System control performance	Diagnostic performance of practiced faults
	Diagnostic performance of novel faults	Diagnostic performance of novel faults	

Author's personal copy

they designed for students of a police university on the basis of a "freezing" technique coupled with debriefing. This approach calls for randomly stopping, or freezing, a simulated task, posing the participant questions about the three levels of SA (Endsley, 1995b), then debriefing the individual. Because little is known about either SA training or the attendant retention intervals, we also investigate both matters in this study.

Explanations for the effectiveness of SA training are that it improves the ability to make timely and effective decisions and to anticipate events (Saus et al., 2006). In addition, SA training facilitates the construction of mental models that improve the understanding of both the importance of various situations and the resources relating to them (Endsley & Robertson, 2000).

1.1.3. Drill and practice (D&P)

Lastly, practice can aid attention performance by proceduralizing or automating a task in order to free resources for another task (Wickens & McCarley, 2008). As individuals continuously practice a task, gradual improvement in time-sharing performance and divided attention has been observed. As soon as one task has been automatized, attentional resources can be applied to other tasks. These changes are ascribed to two processes. First, interference between tasks depends on the demands of the tasks for a limited supply of mental resources. Second, the resource demand of a task decreases with practice until resource-free automaticity is reached (Wickens & McCarley, 2008). We therefore conducted drill and practice (D&P) training as a third approach for process control. D&P facilitates learning through rehearsal of a task in order to achieve a desired level of proficiency (Cannon-Bowers, Rhodenizer, Salas, & Bowers, 1998). Similarly in the EST approach, novice learners are provided with strategies instead of trying to cope with the task on their own. In D&P, learners are provided with a clear strategy to cope with the task and guided in a step-by-step manner through the steps of the task. Thereby, a learner's attention is guided to the accurate execution of the task steps instead of having the learner divide attention between finding a strategy and performing the task at the same time. As with EST, D&P is expected to reduce the learner's mental workload, especially in initial learning.

So far, D&P has been applied mainly for nondynamic and relatively easy cognitive tasks (Shute & Gawlick, 1995). Research has shown practice to be effective at improving accuracy and speed of performance on perceptual, motor, and cognitive tasks (Ericsson, Krampe, & Tesch-Romer, 1993). Practice has also proved to be a more complex training approach than originally thought (Cannon-Bowers et al., 1998; Schmidt & Bjork, 1992). D&P has been especially effective as a method for training people to diagnose practiced fault states (Kluge & Burkolter, 2008; Shute & Gawlick, 1995), though procedural skills demand more practice time than psychomotor skills do (Ginzburg & Dar-El, 2000). Moreover, studies involving initial training and refresher training have shown D&P to support skill retention through the repetition of training exercises (Hagman & Rose, 1983; Schendel & Hagman, 1982; Shute & Gawlick, 1995).

The effectiveness of D&P is attributed primarily to the method's acceleration of component processes and to concomitant restructuring in the use of working memory. It is also attributed to the reduction of the load on working memory that is essential for carrying out cognitive processes (Carlson, Sullivan, & Schneider, 1989).

1.2. The present study

We empirically evaluated the EST, EST/SA, and D&P approaches for their effectiveness over a retention interval of several weeks and in practiced and novel fault states. Because D&P had already been successfully employed (Kluge & Burkolter, 2008), it served as a baseline against which to assess EST and EST/SA. We conducted EST on two main subtasks of process control—(a) system control and stabilization and (b) diagnostic performance—shifting the emphasis between them. This procedure, known as "double manipulation," has been shown to optimize EST (Gopher et al., 1989). In the EST/SA training, EST was supplemented by SA training based on the freezing-and-debriefing technique (see Saus et al., 2006). We used a simulated multitask environment that corresponded to a process control environment. The simulation thus involved the two main subtasks of process control named above.

We derived four assumptions from the literature and studies cited in this study (see Sections 1.1.1–1.1.3). First, the EST/SA group would be more successful than the other groups at developing and maintaining SA, since they receive SA training based on the freezing-and-debriefing technique described above, aimed directly at improving SA.

Second, EST and EST/SA would be more effective than D&P at developing participants' performance on system control. We suggest that, as EST enhances attention-management strategies, participants will be better able to pay careful attention to the state of the parameters and the detection of deviances from the target range, as needed for good system control and stabilization performance.

Third, D&P would be more effective than the two other training methods for developing participants' performance on diagnosing familiar fault states. This assumption is based on research which has shown that D&P supported skill retention, especially for practiced fault states.

Fourth, EST and EST/SA would be more effective for developing participants' performance on diagnosing novel fault states than on diagnosing practiced ones. This assumption is backed first by empirical studies employing EST, which was shown to improve the transfer of skill to new and changed tasks. Second, the training of attention-management strategies is assumed to improve responsiveness to unexpected, novel fault states. Regarding SA training, we suggest that SA training has the potential to improve the diagnostic performance of unfamiliar fault states, for operators with sound SA might detect and understand abnormal situations earlier than they otherwise would. Moreover, they are trained in predicting future states that can evolve out of abnormal states.

2. Method

2.1. Design

A $3 \times 3 \times 2$ mixed factorial design was employed. Training as a between-participants variable varied at three levels (EST, EST/SA, and D&P). Time of measurement as a within-participant variable was taken in three separate testing sessions ($test_0$, $test_{2w}$, and $test_{6w}$). Fault type as a within-participants variable varied at two levels (practiced and novel faults).

2.2. Participants

Forty-eight university students (four female) participated in the experimental study. They were all enrolled in a program leading to a Bachelor of Science degree in a technical field of study (aeronautics, engineering, or electrical engineering) at universities of applied sciences in the Greater Zurich Area, Switzerland. The students were paid 100 CHF (approximately US $90.00) for participation in all three parts of the experiment. Participants were randomly allocated to the three training methods at each location. Forty students (83.3%; four female) completed all three parts of the experiment. Training groups did not differ significantly regarding

Author's personal copy

drop-outs ($H(2)$=1.336, $p > 0.05$). The mean age of the participants was 24.7 years ($SD = 4.0$). There were no significant differences between the mean ages of the training groups ($F(2, 39) = 0.874$, $p > 0.05$).

2.3. Simulated process control environment

In an introductory training module and subsequent testing sessions, we used a computerized process control task simulated by the Cabin Air Management System (CAMS; for details, see Sauer, Wastell, & Hockey, 2000). CAMS models a life support system onboard a spacecraft. Five parameters (O_2, CO_2, cabin pressure, temperature, and humidity) are kept in a predefined zone by automatic controllers. The operator's task is to intervene if necessary. This individual must complete two primary tasks (system control and fault diagnosis, see Sections 2.5.3 and 2.5.4) and two secondary tasks (prospective memory and reaction time). CAMS records the actions carried out by the operator.

2.4. Training methods

Three different types of training were given: EST (13 participants), combined EST and SA training (14 participants), and D&P (13 participants). As is often the case with training that involves complex systems such as refinery simulators, training was carried out in small groups typically consisting of four to six participants supervised by one instructor in order to ensure sound supervision and effective learning (Kriedemann, 2008).

All training had the same general introduction to CAMS and focused on the same five system faults. Training material given to all training groups consisted of an illustration of CAMS, its components, and controls (with notations and translations); a CAMS manual; and an instruction manual. The CAMS manual described the main components, systems and controllers of CAMS, the tasks of the participants and 16 different system faults. For every system fault, a description of the fault was given, the symptoms were described and the intervention steps (system control and fault-finding) were depicted. The instruction manual (see Fig. 1) was designed to guide the participant through the training of the five system faults. The instruction manual was based on the CAMS manual and included a screenshot of CAMS during the fault state and descriptions of symptoms and intervention procedures for fault diagnosis and repair and for control and stabilization of the system. The instructions and the number of exercises per system fault differed from one training group to the next, but the duration of training was the same for all groups. Whereas the EST and EST/SA group worked with the same instruction manual, the D&P group received a different instruction manual (see Sections 2.4.1 and 2.4.2).

The first training block (see also Fig. 3) was a general introduction to CAMS and the corresponding manual. The participants received 10 min of multimedia instruction about CAMS, its features, and the primary and secondary tasks involved in the experiment. Participants followed the multimedia instructions individually on their computers using earphones. They were then given a few minutes to explore the system on their own (e.g. looking at the system components, trying out controllers). A short presentation and an exercise introducing the CAMS manual followed. The presentation was given by the instructor and aimed to prepare participants to use the manual by describing its content, structure, and function. In the exercise, participants were asked to find a certain system fault ("On which page can the system fault 'vent stuck on' be found?"), and to describe the system fault and its symptoms. Answers were discussed with the instructor.

The second training block introduced all participants to five system faults: (a) a leak in an oxygen (O_2) valve, (b) a cooler set point failure, (c) a block in a mixer valve, (d) a carbon dioxide (CO_2) set point failure, and (e) a nitrogen set point failure. Selection of system faults was based on a hierarchical task and an analysis of subgoal templates (see Burkolter, Kluge, Schüler, Sauer, & Ritzmann, 2007). System faults were randomly allocated over parameter, type of faults, and the difficulty of the repair procedure. The participants

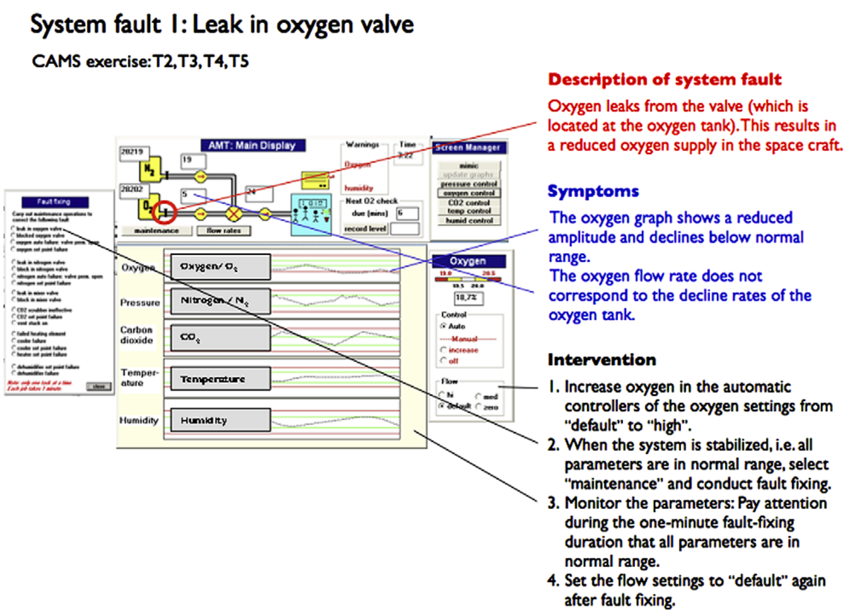

Fig. 1. Instruction manual for the D&P group depicting one of five system fault descriptions (translated from German). On the left side, a screenshot of the simulation CAMS during the fault state is displayed. On the right side, the system fault, the symptoms, and the steps of intervention are described. Faults, symptoms and intervention steps are indicated on the screenshot with straight lines.

Author's personal copy

practiced the five systems faults by following the instruction manual. Each first exercise of a system fault was conducted using a pause button that is provided in CAMS. Thus, the participants had the possibility to pause CAMS and refer to the instruction manual for further intervention steps.

Participants worked individually on their own computers by following the instruction manual and the instructions provided by the instructor. The participants were given the possibility to ask questions at any time during the training.

2.4.1. Emphasis shift training (EST)

At the beginning of the second training block, the instructor explained the principles of EST by the means of a presentation. The two main tasks of CAMS, the principle of changing the emphasis between the two, and the idea of EST (to learn to manage more than one task at the same time) were pointed out. Then, a 10-min introduction to system control and stabilization was provided along with a corresponding exercise (increasing three different parameters by adjusting controls and observing what happens). Thereafter, exercises on the five system faults to be trained followed. The instructor trained the EST group by changing the emphasis on the two main tasks of process control (system control and fault-finding). In each exercise, one of these main tasks was alternately emphasized. That is, participants practiced the actions of only one main task and did not have to execute the other interventions. The instruction manual of the EST group indicated the main task to be emphasized, and the proper steps were marked with a red arrow and boldfaced letters. Information on the other task appeared in light coloring (see Fig. 2). In all other respects, the instruction manual was identical to the one the D&P group received. The members of the EST group practiced each system fault (SF) three times: twice one fault separately and once together ($2 \times SF_1$, $2 \times SF_2$, $1 \times SF_{1\ \text{and}\ 2}$ together; $2 \times SF_3$, $2 \times SF_4$, $2 \times SF_5$, $1 \times SF_{3,\ 4,\ 5}$ together).

2.4.2. EST combined with SA training (EST/SA)

The EST/SA group and EST group received the same instruction manual and the same exercises with changing emphasis. Additionally, the EST/SA group was given an SA training.

As in the EST group, the instructor explained the principles of EST by means of a presentation. Then, a 10-min introduction to system control and stabilization was provided along with a corresponding exercise. This was followed by exercises on the first system fault to be trained. Participants were trained by changing the emphasis on the two main tasks of process control. Then, in contrast to EST, a brief presentation introducing the concept of SA was given by the instructor. The concept of SA, the three levels of SA (Endsley, 1995b), and its relevance for CAMS performance were explained to the participants. From the third system fault on, SA was practiced through the freezing technique with debriefing (as described by Saus et al., 2006). The simulation was frozen one time during each exercise, and participants received open questions about two parameters based on Endsley's (1995b) concept of SA. The first item concerned perception of the situation (e.g. "How is the status of temperature?"); the second item concerned comprehension and future actions ("What does this mean? What needs to be done now?"). After completing the short questionnaire, the participants discussed responses with the instructor. The points at which the simulation was frozen were distributed over the beginning, middle, and end of a system fault. The questions took all parameters into account.

2.4.3. Drill and practice (D&P)

The D&P group extensively practiced system control and fault-finding by repeating several different drills. The participants were told to follow the steps of intervention closely as described in the instruction manual (see Fig. 1), which explained the various steps of system control and stabilization, fault diagnosis, and repair. For each system fault, they were first given 3 min to read about it and to memorize the intervention steps. They then practiced

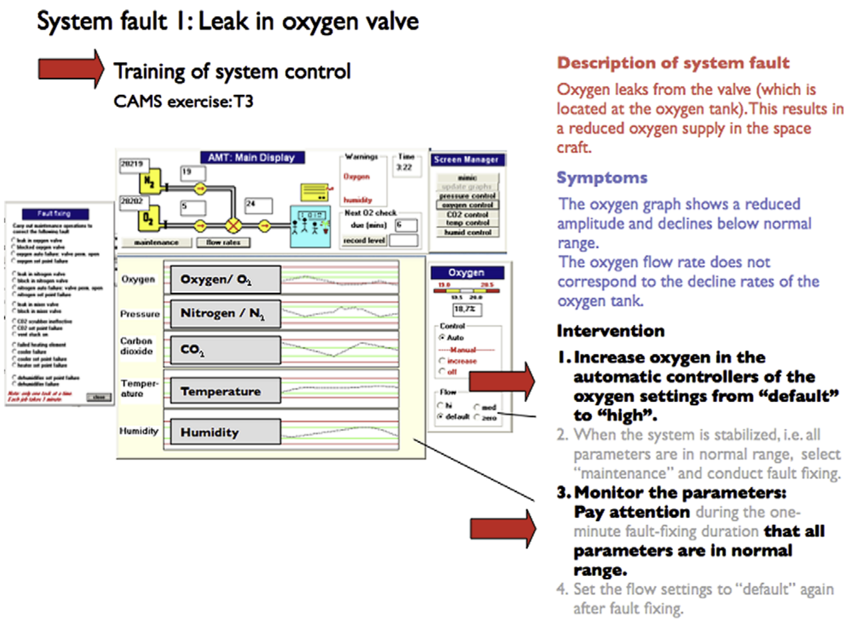

Fig. 2. Instruction manual for the EST and EST/SA group depicting one of five system faults (translated from German). This is an exercise emphasizing the training of system control, which is indicated at the top left for the participants. The steps of intervention that concern system control are in boldface and marked with a red arrow. Information on the other task, i.e. fault-finding (as well as description and symptoms), appear in light coloring.

Author's personal copy

each system fault with its intervention steps a total of five times—four times each fault separately, then once together with at least one other system fault (SF) ($4 \times SF_1$, $4 \times SF_2$, $1 \times SF_{1 \text{ and } 2}$ together; $4 \times SF_3$, $4 \times SF_4$, $4 \times SF_5$, $1 \times SF_{3,\,4,\,5}$ together).

2.5. Measures

2.5.1. Reactions to training

After the training module the participants rated five reactions to the training they had just experienced. The first three—mental effort, anxiety, and fatigue—were each rated in response to a single item each. The item for anxiety, for example, read "How do you feel right now?" and was rated on a scale from 0 (e.g. calm) to 100 (e.g. tense) (see Sauer, Wastell. & Hockey, 2000, for details). The last two reactions were motivation and self-efficacy. They were rated on a six-point scale ranging from 1 (I totally agree) to 6 (I do not agree at all). The participants' responses relating to post-training motivation was elicited by the following four items: "The task appealed to me." "I would like to participate again in a training like this." "I was motivated to accomplish the CAMS task." "I find CAMS interesting and fascinating" (Cronbach's Alpha = 0.86). The degree of self-efficacy that trainees felt about understanding and controlling CAMS was measured through their responses to four items (e.g. "I feel up to the tasks of CAMS." (Cronbach's Alpha = 0.87; see Kluge, 2008, for details).

2.5.2. Situation awareness

To check whether SA training was improving SA, we measured the latter index by means of the SA Control Room Inventory (Hogg et al., 1995), which was adapted for application to CAMS. The inventory was developed for use in process control research based on the Situation Awareness Global Assessment Technique (Endsley, 1995a). During the testing sessions, the simulation was frozen six times at irregular intervals. The participants were instructed to switch off their screens so that they could not see the current system state. Questions about the status of the system were then presented. In accordance with Endsley's (1995b) concept of SA, one item concerned the perception of the current situation ("How is the status of humidity?"). The response alternatives were "below normal range," "within normal range," and "above normal range." Another item covered comprehension of the current situation and prediction of future status ("How do you think the course of humidity will develop over the next 10 s? Provided that no intervention to the system is undertaken, the parameter will ... "decline," "remain stable," or "increase" in 10 s."). SA was measured once in the first CAMS test run and once in the second CAMS test run. There were twelve questions in total (six measurements with each two items). The responses were then compared to logged CAMS data to determine whether the responses by the participants were correct (for this approach see Hogg et al., 1995, p. 2411).

2.5.3. System control failures

One of the primary tasks of the operator was to maintain five key parameters within normal range. If one or more of the key parameters departed from normal range, the operator needed to intervene by adjusting automatic controllers or adapting manual control. The duration of the parameters' deviation from normal range was measured in seconds and converted into percentages.

2.5.4. Fault-finding

The other primary task concerned fault diagnosis and repair with the assistance of the maintenance facility. In the event of a system fault, the operator had to identify the cause by carrying out appropriate tests. There were two measures of diagnostic performance: the percentage of incorrect diagnoses (diagnostic accuracy) and the number of seconds the operator needed in order to identify the fault correctly (diagnostic speed).

2.5.5. Knowledge tests

Knowledge was assessed with an adapted version of two existing knowledge tests on CAMS (see Sauer, Burkolter, Kluge, Ritzmann, & Schüler, 2008). Structural knowledge was measured with a method described by Meyer (2008). Findings concerning knowledge are reported in a separate article (Burkolter, Meyer, Kluge, & Sauer, in press).

2.6. Procedure

Fig. 3 summarizes the experimental procedure. There were three parts. The first consisted of questionnaires and a training module (about 4¼ h) followed immediately by a 45-min testing session ($test_0$). The second part was a 45-min testing session two weeks after the training module ($test_{2w}$). The third part was a 45-min testing session six weeks after the training module ($test_{6w}$). Because of organizational constraints stemming from the differences between the schedules of university terms, the retention interval between the testing sessions could not be identically long (i.e. two weeks between $test_0$ and $test_{2w}$ and four weeks between $test_{2w}$ and $test_{6w}$).

Upon arrival at the experimental facility, all participants spent 35–45 min completing questionnaires on cognitive ability, cognitive flexibility, personality, and motivation. This pretraining testing did not include a testing session on the CAMS task to rule out initial differences of the training groups in process control performance. However, CAMS is an artificial task which is not commercially available, cannot be retrieved from the internet and is not known outside the scientific community. Although CAMS is situated in the context of spaceflight, it does not correspond directly to real-world physical principles, but has its own rules and interconnections. Therefore, previous knowledge on spaceflight and related knowledge is unlikely to be very useful for CAMS performance, and thus the likelihood of initial differences of training groups are assumed to be low. However, we controlled for differences between training groups regarding cognitive and personality variables.

The training module was equally long for all three training groups. It lasted approximately 3¼ h (including one 5-min and one 20-min break). The number of training exercises varied, however. The D&P group performed five exercises per system fault, for the core idea of D&P is to provide a good deal practice on the task. During the SA training of the EST/SA group, the members of the EST group were given a comparable cognitive task. They heard a talk (about recruitment criteria for astronauts; see Sauer et al., 2008), viewed part of a documentary, and like the members of the EST/SA group, answered questions about the material to which they had just been exposed (about 35 min). $Test_0$ covered all five practiced fault states in addition to fault states that the participants had not previously encountered (block in nitrogen valve and dehumidifier set point failure).

$Test_{2w}$ and $test_{6w}$ were identical for all participants. After a brief introduction to refresh knowledge on the experimental task, participants worked with CAMS and were tested for SA during either the first or second part of the test (see Section 2.5.2). $Test_{2w}$ was followed by the knowledge tests, which took 30–40 min. The second and third testing sessions each included three familiar and three novel fault states. For an overview of the system faults employed in training and testing sessions, see Table 2. Participants were given no advance information about the order and time of appearance of faults. The CAMS manual was available to participants during the testing sessions. After the final testing session, participants were debriefed about the three training methods

Author's personal copy

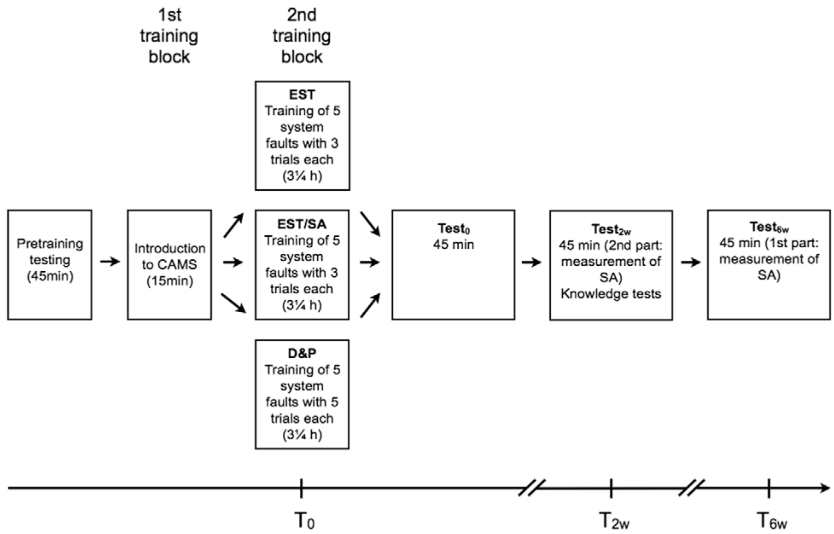

Fig. 3. The experimental procedure. EST: emphasis shift training; EST/SA: EST and situation awareness (SA) training; D&P: drill and practice. Test_0: test immediately after training; Test_{2w}: test two weeks after training; Test_{6w}: test six weeks after training.

and given the opportunity to ask questions about the design of the experiment.

During the first part of the experiment in which the training took place, it was assured that participants in the different training groups could not interact or learn from each other. The training sessions were conducted in different rooms or on different days. Thus, in the particular part of the experiment in which the actual training intervention took place, participants could not interact with each other. In the two weeks between the first and second as well as the four weeks between the second and third sessions, however, participants could have had the opportunity to interact. Unfortunately, this could not have been prevented, since participants were enrolled at the same universities and (partly) attending the same courses. The latter is associated with the fact that we conducted the experiment at the particular universities of the participants instead of inviting them to a lab. However, it was assured that none of the instruction material including the CAMS program was available to the participants outside the experimental sessions by collecting all material at the end of the sessions. Moreover, there is no single best "solution" to the CAMS task that could have been shared by participants, but during testing sessions, the CAMS manual was available to all participants.

Table 2
Overview of system faults employed in training and testing sessions.

Fault type	Training	Test_0	Test_{2w}	Test_{6w}
Practiced faults				
Leak in oxygen (O_2) valve	×	×	×	
Cooler set point failure	×	×	×	×
Block in mixer valve	×	×	×	
CO_2 set point failure	×	×		×
Nitrogen set point failure	×	×		×
Novel faults				
Block in nitrogen valve		×		
Dehumidifier set point failure		×		
Nitrogen valve permanently open			×	
CO_2 scrubber ineffective			×	
Oxygen valve permanently open			×	
Block in oxygen (O_2) valve				×
Cooler failure				×
Leak in mixer valve				×

2.7. Assessment of control variables

We controlled for cognitive ability, cognitive flexibility, conscientiousness, and pretraining motivation, which are seen as relevant for training situations (see Colquitt, LePine, & Noe, 2000). These variables were employed to control for possible "unhappy randomization" (Mohr, 1995). Cognitive ability was assessed with the Wonderlic Personnel Test (Wonderlic Inc., 2002) cognitive flexibility, with the Cognitive Flexibility Inventory (Spiro, Feltovich, & Coulson, 1996). Big-Five Markers (Saucier, 1994) were employed to measure conscientiousness as a personality trait. Pretraining motivation was assessed with one item ("Please indicate how motivated you are to participate in this training? 0%, 20%, 40%, 60%, 80%, 100%."). As Table 3 shows, participants in the three groups did not differ significantly with regard to control variables. Thus, there was no unhappy randomization.

3. Results

We performed mixed ANOVAs. If the assumption of sphericity was violated, we corrected degrees of freedom by using Greenhouse-Geisser estimates of sphericity. We drew planned contrasts in specific ways (see Loftus, 1996). The first contrast was between the D&P group and the EST/SA group, the purpose was to compare the two groups that had had only exercises and SA training and no further task. We focused the second contrast on the EST and EST/SA groups so as to ascertain whether the additional SA training enhanced process control performance beyond the effects of EST. Interaction effects were broken down into interaction contrasts, as proposed by Gamst, Meyers, and Guarino (2008).

Wickens (1998) states that low sample size and high variance may increase the probability of a Type II error and that, ergonomically speaking, Type II statistical errors can be as important as Type I errors. To avoid Type II statistical errors, we also report *p*-values at the 10% level.

3.1. Reactions to training

Training groups differed significantly in their ratings of mental effort, anxiety, and fatigue (see Table 4). Planned contrasts of effort

Author's personal copy

revealed that the D&P group differed significantly from the EST/SA group ($p < 0.05$), with the D&P group investing the most effort in the task. In terms of anxiety and fatigue, planned contrasts revealed that the EST group was less tense and tired than the EST/SA group ($p < 0.1$). However, participants did not differ significantly on either post-training motivation or self-efficacy ratings.

3.2. Situation awareness

Surprisingly, the participants of the three training groups did not differ significantly on their SA performance tested at T_{2w} and T_{6w} ($F(2, 35) = 0.097$, $p > 0.10$, $\eta_p^2 = 0.00$). Neither a significant main effect of time ($F(1, 35) = 0.463$, $p > 0.10$, $\eta_p^2 = 0.01$) nor a significant interaction effect of group and time was found ($F(2, 35) = 0.524$, $p > 0.10$, $\eta_p^2 = 0.03$). Thus, our first assumption was not supported by the data.

3.3. System control failures

An inspection of descriptive statistics (see Table 5) suggests that there was a main effect of training and that the EST group performed better in system control than the other two groups did. However, the three-way mixed ANOVA failed to support this interpretation ($F(2, 37) = 1.49$, $p > 0.10$, $\eta_p^2 = 0.07$). In contradiction of the second assumption there was no significant main effect of training. A significant interaction effect between fault type and training group was observed, however, ($F(2, 37) = 3.37$, $p < 0.05$, $\eta_p^2 = 0.15$). This interaction indicates that training groups differed in their performance depending on the type of fault (either practiced or novel). As depicted in Fig. 4, the D&P and EST groups performed better than the EST/SA group during practiced faults, and the EST/SA group performed better during novel faults. Interaction contrasts showed a significant interaction for the D&P group and EST/SA group ($F(1, 25) = 8.992$, $p < 0.01$, $\eta_p^2 = 0.27$) and for the EST and EST/SA group ($F(1, 25) = 2.895$, $p = 0.10$, $\eta_p^2 = 0.1$). Analysis also revealed a significant main effect of both time ($F(1.73, 37) = 35.08$, $p < 0.01$, $\eta_p^2 = 0.49$) and fault type ($F(1, 37) = 5.861$, $p < 0.05$, $\eta_p^2 = 0.14$). Contrasts regarding the main effect of time showed that performance differed significantly between T_0 and T_{2w} and between T_{2w} and T_{6w} ($p < 0.001$), with the best performance occurring at T_{2w}.

3.4. Diagnostic performance

A three-way mixed ANOVA with diagnostic accuracy was performed (see Table 6). Analysis revealed no significant effect of training group on performance ($F(2, 37) = 0.795$, $p > 0.10$, $\eta_p^2 = 0.04$). Confirming the assumptions, however, there was a significant interaction effect between fault type and training ($F(2, 37) = 2.72$, $p < 0.10$, $\eta_p^2 = 0.13$), indicating that the performance of the training groups differed in fault type. The interaction graph (see Fig. 5) displays that D&P resulted in better performance on practiced faults than on novel faults, whereas EST/SA resulted in better performance on novel faults than on practiced faults. Interaction contrasts showed a significant interaction for the D&P group and EST/SA group ($F(1, 25) = 4.072$, $p < 0.10$, $\eta_p^2 = 0.14$), confirming the assumption that EST/SA aided diagnosis of novel faults. A significant main effect of time was found as well ($F(2, 74) = 6.45$, $p < 0.01$, $\eta_p^2 = 0.15$). Contrasts showed that performance at T_0 and T_{2w} and at T_{2w} and T_{6w} differed significantly ($p < 0.01$), with the poorest performance generally occurring at T_{2w}. There was no main effect of fault type ($F(1, 37) = 0.06$, $p > 0.10$, $\eta_p^2 = 0.00$).

On the second measure of diagnostic performance (diagnostic speed), the results of the three-way mixed ANOVA resembled those relating to diagnostic accuracy (see Table 6). We observed no significant main effect of training ($F(2, 37) = 0.51$, $p > 0.10$, $\eta_p^2 = 0.03$), but, as with diagnostic accuracy, the interaction between fault type and group was significant ($F(2, 74) = 3.87$, $p < 0.05$, $\eta_p^2 = 0.17$). This interaction effect indicated that training groups differed significantly in diagnostic speed regarding fault

Table 3
Descriptive statistics (*M*, *SD*) on control variables as a function of training group.

Control variables	Drill and practice	EST	EST/SA	*F*/*p*
Cognitive ability (0–50)	25.71 (4.68)	26.25 (6.26)	24.88 (6.04)	$F(2, 45) = 0.232$, $p > 0.05$
Cognitive flexibility (–4 to +4)	1.00 (1.22)	0.84 (0.90)	0.53 (1.07)	$F(2, 46) = 0.814$, $p > 0.05$
Conscientiousness (1–9)	6.84 (0.95)	6.52 (1.00)	6.78 (1.30)	$F(2, 46) = 0.360$, $p > 0.05$
Pretraining motivation (%)	80.00 (13.59)	80.0 (16.90)	76.47 (19.02)	$F(2, 45) = 0.236$, $p > 0.05$

Table 4
Descriptive statistics on reactions to training as a function of training group.

Reactions to training	Drill and practice *M* (*SD*)	EST *M* (*SD*)	EST/SA *M* (*SD*)	*F*/*p*
Effort (0–100)	68.93 (20.0)	48.00 (21.17)	50.83 (26.13)	$F(2, 47) = 3.644$, $p < 0.05$, $\eta_p^2 = 0.14$
Anxiety (0–100)	47.79 (23.73)	27.59 (19.53)	40.56 (23.47)	$F(2, 47) = 3.198$, $p < 0.1$, $\eta_p^2 = 0.12$
Fatigue (0–100)	68.96 (26.17)	47.06 (26.47)	64.17 (26.70)	$F(2, 47) = 2.936$, $p < 0.1$, $\eta_p^2 = 0.12$
Motivation (1–6)	4.13 (0.81)	4.21 (1.04)	3.68 (1.27)	$F(2, 45) = 1.147$, $p > 0.1$, $\eta_p^2 = 0.05$
Self-efficacy (1–6)	2.73 (0.94)	2.75 (0.94)	3.19 (1.13)	$F(2, 47) = 1.120$, $p > 0.1$, $\eta_p^2 = 0.05$

Table 5
System control failures (in percentages) as a function of training and fault type (SD in parentheses).

Fault type	Drill and practice			EST			EST/SA		
	T_0	T_{2w}	T_{6w}	T_0	T_{2w}	T_{6w}	T_0	T_{2w}	T_{6w}
Practiced	11.6 (9.0)	3.5 (2.3)	8.6 (5.5)	9.5 (8.0)	4.0 (6.0)	6.4 (4.0)	17.6 (14.1)	5.1 (5.7)	9.8 (6.1)
Novel	8.8 (6.9)	7.4 (6.3)	17.5 (3.6)	6.6 (2.3)	4.3 (3.1)	15.5 (5.3)	8.4 (4.6)	6.2 (4.9)	15.7 (6.9)

Author's personal copy

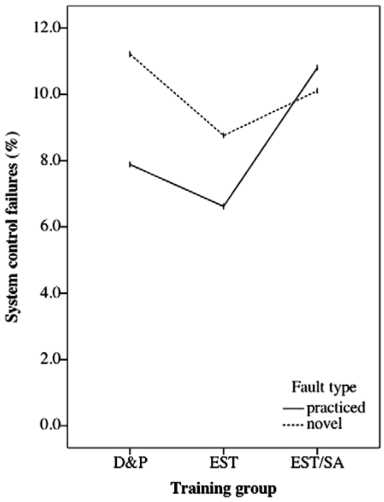

Fig. 4. Interaction between training group and fault type for system control performance.

type (Fig. 5). Interaction contrasts showed a significant interaction for the D&P group and EST/SA group ($F(1, 25) = 7.286$, $p < 0.05$, $\eta_p^2 = 0.23$). Moreover, a significant main effect of time ($F(2, 74) = 15.3$, $p < 0.001$, $\eta_p^2 = 0.29$) and fault type was observed ($F(1, 74) = 6.66$, $p < 0.05$, $\eta_p^2 = 0.15$). Contrasts for main effect of time revealed that performance at T_0 and T_{2w} and at T_{2w} and T_{6w} differed significantly ($p < 0.001$), with the poorest performance generally occurring at T_{2w}.

4. Discussion

Using the training methods EST, EST/SA, and D&P, we examined attention skills and process control performance in familiar and nonfamiliar situations over a retention interval of several weeks. We aimed to support novices learning a highly complex and demanding task by providing them with attention-management strategies in order to reduce their mental workload. D&P was successful at enhancing diagnostic performance on familiar system faults, and EST/SA training supported the diagnosis of novel system faults. All in all, EST and EST/SA did not support system control performance as strongly as we had assumed they would.

D&P proved effective at increasing the speed and accuracy with which participants found familiar fault states. Thus, it seems that D&P was successful at providing trainees with clear strategies to cope with the high demands of the task. Participants were guided through the steps of a task so that attentional resources could be applied to learning instead of finding a strategy to cope with the task on their own. The finding that D&P improved diagnostic performance on practiced fault states confirms earlier results of research involving the CAMS task in a comparable experimental setting (Kluge & Burkolter, 2008). In that experiment, error training, procedure-based training with error-relevant heuristics, and D&P were compared for their effectiveness at enhancing process control performance. D&P emerged clearly as the most effective method for developing the skill of diagnosing familiar fault states, even after retention intervals of 9 and 13 weeks. These findings suggest that the use of D&P was broadened from nondynamic and rather easy cognitive tasks (Shute & Gawlick, 1995) to more dynamic and highly complex cognitive tasks. However, the findings also show that the effectiveness of D&P was confined to familiar fault states. In other words, participants were prepared to apply acquired skills but limited in their ability to adapt skills to new situations. This finding supports the contention that training methods concentrating on teaching procedures rather than teaching knowledge have a restricted range of transfer (Hockey, Sauer, & Wastell, 2007). Lastly, participants in the D&P group were not less motivated than participants in the other two groups, although D&P requires repetitive work on a task. This result might be of special interest to instructors.

Possible explanations why the experiment did not confirm some of the assumptions regarding EST are discussed. The aim of the EST was to enhance attention management by introducing emphasis changes on components of a task. Thus, participants were provided with strategies to reduce mental workload and improve performance. The selection of EST as a training approach for process control was carefully based on empirical evidence, theoretical considerations, and its successful use in different contexts (Gopher, 2007; Gopher et al., 1989). Even so, our experiment showed EST to have only slightly positive interaction effects and did not detect any clear advantage for EST. We note, however, that the present study was more complex in task and design than previous ones on this subject. First, we used a complex and dynamic process control environment that entailed fewer psychomotor and higher cognitive demands than have tasks previously posed in like settings. CAMS involves task management activities such as monitoring to detect deviance and changes, actions to stabilize the system, retrieving information, diagnosing, planning, forming rules, and evaluating actions (Burkolter et al., 2007; Ormerod, Richardson, & Shepherd, 1998). The Space Fortress game employed by Gopher et al. (1989) involved manual control and discrete, precise motor responses, visual scanning and monitoring, memory requirements, and decision-making. Similarities between these tasks may have been overestimated. Second, we used a multifactorial design that included three different points in time. Previous findings concerned positive effects on performance at the end of training, not skill retention over several weeks. All in all, the "take-the-best" application and transfer of EST to a different field seemed to entail unexpected difficulties, and it was not as successful in this study as in previous ones (see Gopher, 2007). Methodologically, it could be argued for a larger sample, for small sample sizes increase the likelihood of incorrectly concluding that there is no statistical difference (Cook, Campbell, & Peracchio, 1990). However, the statistical significance one might gain with a larger sample will not necessarily improve the experiment's practical significance.

Table 6
Diagnostic performance as a function of training and fault type (SD in parentheses).

Fault type	D&P			EST			EST/SA		
	T_0	T_{2w}	T_{6w}	T_0	T_{2w}	T_{6w}	T_0	T_{2w}	T_{6w}
Diagnostic errors (%)									
Practiced	21.5 (12.8)	51.3 (17.3)	33.3 (19.2)	40.0 (29.4)	59.0 (30.9)	41.0 (30.9)	50.0 (35.7)	57.1 (27.5)	52.4 (31.3)
Novel	46.2 (24.7)	51.3 (35.0)	46.2 (32.0)	38.5 (30.0)	51.3 (25.9)	38.5 (18.5)	53.6 (23.7)	52.4 (38.6)	35.7 (33.2)
Diagnostic speed (s)									
Practiced	223.4 (58.9)	307.8 (59.4)	274.4 (70.5)	241.7 (87.4)	347.5 (83.4)	290.8 (92.9)	292.0 (89.5)	332.1 (76.6)	327.1 (94.0)
Novel	300.6 (74.5)	336.5 (84.1)	358.3 (80.7)	271.3 (65.4)	310.3 (89.0)	330.4 (70.0)	298.4 (80.0)	337.6 (115.5)	319.9 (95.5)

Author's personal copy

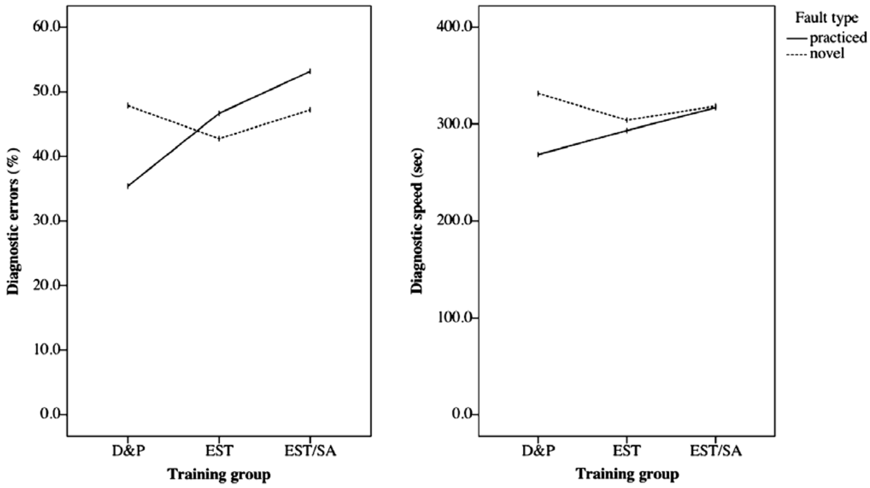

Fig. 5. Interaction between training group and fault type for diagnostic errors (left) and speed (right).

The EST/SA approach showed positive effects for fault-finding in novel situations, but it was not as effective as we assumed it would be for system control and SA. By combining EST and SA training, we aimed both to enhance SA and performance and to support SA by improving attention management. There are reasons to believe that the SA approach might have interfered with EST, an effect especially challenging to novices. In EST, procedures are taught through rule-based instruction, which only implicitly communicates properties of the system (Rasmussen, 1990). To answer the questions relating to SA, however, participants were required to anticipate system states. That is, they had to shift from a rule-based to a knowledge-based level. The combination of training methods that require cognitive processing at different levels may therefore have been too challenging for novices. It seems that the expected positive effects of EST and positive effects of SA were neutralized rather than compounded, at least at this stage of learning. One could consider delaying EST/SA training (see Schneider, 1985) until, say, D&P has helped participants firmly establish a procedure for the main tasks. This sequencing could enhance attention performance by automating a task and thereby freeing attention resources for accomplishing another task (Wickens & McCarley, 2008).

Regarding the measurement of SA, Vidulich (2003) assumes a rich interplay between specific memory of the current situation and a skilled individual's long-term memory. Long-term working memory is assumed to serve as the basis to answer questions during the freeze of the simulation. Possibly, participants had not yet developed enough expertise as a basis to answer the SA questions. With respect to methodological issues, results showed that the SA measurement did not significantly differ from one training group to the next. This finding might be an indication of validity issues with the SA measure we employed—based on the SA Control Room Inventory (Hogg et al., 1995). We note that all the participants scored relatively low regardless of what training group they were in, suggesting a floor effect. Further research employing the novel SA measure is needed to test this possibility.

Surprisingly, performance on system control decreased from the second testing session (two weeks after training) to the third (six weeks after training). By contrast, diagnostic performance was poorer at the second testing session and improved at the third one. These results concerning skill retention after training were similar for all training groups which is an interesting finding, especially because previous research on skill retention has focused mainly on single tasks (see Arthur, Bennett, Stanush, & McNelly, 1998). By contrast, the two tasks in our study—system control and fault-finding—had to be accomplished at the same time. Therefore, findings from single-task studies (e.g. Arthur et al., 1998) might not directly apply for transfer to dual tasks. As abilities of operators in the two main tasks of process control, system control and diagnostic performance are independent (Landeweerd, 1979), it might be possible that skill decay of the two tasks is also dissimilar. This assumption is supported by findings from two experiments with the same simulated process control task (Kluge & Burkolter, 2008; Burkolter, Kluge, & Brand, 2009), in which a similar pattern of results was observed. Whereas system control performance decreased from a first (9 weeks after training and directly after training, respectively) to a second testing session (13 weeks after training and one week after training, respectively), diagnostic performance increased from the first to the second testing session. While these effects have not yet been investigated in detail, we speculate whether participants might have concentrated more on the system control task if training had not taken place long ago. However, at later testing points, when participants might have felt that there were shortcomings in remembering the task, they possibly concentrated more on diagnosis, for which they could find specific information not only on intervention but also on the description of system faults and the symptoms in the manual. However, this issue needs further analysis and research.

Some limitations regarding the study procedure should be pointed out. We conducted the study with students, and not with experienced operators working in process control environments. However, we did invite engineering students to participate in our experiment in order to enhance transfer of study results to process control. These students participated voluntarily in the study as an extracurricular activity, which might indicate a high level of motivation. On the other hand, the extensive training might also have implied an additional workload on them. Moreover, participants did not practice the task between the experimental sessions, which might not directly apply to real-world settings, where operators usually work between training sessions and thus also gain experience between training sessions. Moreover, the process control task was new to the participants, therefore entailing initial learning of complex skills. Further research should determine to what extent such results are transferable to further stages of learning and training.

In summary, the present study aimed at contributing to research of workplace training by applying training methods from

Author's personal copy

fields such as aviation, police, and nondynamic, motor tasks (e.g. Gopher et al., 1989; Saus et al., 2006) to a process control environment in order to extend established training research findings to a novel work environment. The problem of limited attention capacities has been discussed with respect to complex tasks (Gopher, 1996) and the limitation they present for SA (Endsley, 1995b; Wickens & McCarley, 2008). We aimed to enhance this work by analyzing the training of attention management and allocation of limited attention resources in process control.

This study confirmed D&P as a promising approach for teaching novices to successfully diagnose familiar fault states in process control for up to several weeks after training. We recommend that further research identifies the conditions under which D&P is also effective with experienced operators. The present experiment might serve as a starting point for detailed investigation intended to gather further evidence about the effectiveness of EST and EST/SA training in process control. That work could be a promising step in the effort to provide operators with effective training designed to improve attention management and address the problem of limited attention capacities.

Acknowledgment

This project has been funded by the Swiss National Science Foundation (No. P001-106354).

References

Arthur, W., Bennett, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57–101.

Burkolter, D., Kluge, A., & Brand, M. (2009). Cognitive and personality variables of operators: Individual characteristics and process control performance. In D. de Waard, F. L. J. Godthelp, & K. A. Brookhuis (Eds.), *Human factors, security and safety* (pp. 77–88). Maastricht: Shaker Publishing.

Burkolter, D., Meyer, B., Kluge, A. & Sauer, J. (in press). Assessment of structural knowledge as a training outcome in process control environments. *Human Factors*.

Burkolter, D., Kluge, A., Schüler, K., Sauer, J., & Ritzmann, S. (2007). Cognitive requirement analysis to derive training models for controlling complex systems. In D. de Waard, G. R. J. Hockey, P. Nickel, & K. A. Brookhuis (Eds.), *Human factors issues in complex system performance* (pp. 475–484). Maastricht: Shaker Publishing.

Cannon-Bowers, J. A., Rhodenizer, L., Salas, E., & Bowers, C. A. (1998). A framework for understanding prepractice conditions and their impact on learning. *Personnel Psychology, 51*, 291–320.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 517–526.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678–707.

Cook, T., Campbell, D., & Peracchio, L. (1990). Quasi experimentation. *Handbook of industrial and organizational psychology* (2nd ed.) (Vol. 1, pp. 491–576). Palo Alto, CA, US: Consulting Psychologists Press.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Endsley, M. R., & Robertson, M. M. (2000). Training for situation awareness in individuals and teams. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 349–365). Mahwah: Lawrence Erlbaum Associates Publishers.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). *Analysis of variance designs. A conceptual and computational approach with SPSS and SAS*. Cambridge: Cambridge University Press.

Ginzburg, S., & Dar-El, E. M. (2000). Skill retention and relearning - a proposed cyclical model. *Journal of Workplace Learning, 12*, 327–332.

Gopher, D. (1996). Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research, 5*, 23–28.

Gopher, D. (2007). Emphasis change as a training protocol for high-demand tasks. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 209–224). New York: Oxford University Press.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*, 387–405.

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

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors, 25*, 199–213.

Hockey, G. R. J., Sauer, J., & Wastell, D. G. (2007). Adaptability of training in simulated process control: Comparison of knowledge- and rule-based guidance under task changes and environmental stress. *Human Factors, 49*, 158–174.

Hogg, D. N., Follesø, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics, 38*, 2394–2413.

Kluge, A. (2008). What you train is what you get? Task requirements and training methods in complex problem-solving. *Computers in Human Behavior, 24*, 284–308.

Kluge, A., & Burkolter, D. (2008). Training principles for process control tasks and innovative approaches to training. In *Joint international symposium of ISSNP 2008/CSEP 2008/ISOFIC 2008: Toward innovative nuclear safety and simulation technology* (Vol. 2, pp. 99–106). Harbin, China: Harbin University Press.

Kluge, A., Sauer, J., Schüler, K., & Burkolter, D. (2009). Designing training for process control simulators: A review of empirical findings and current practices. *Theoretical Issues in Ergonomics Science, 10*, 489–509.

Kriedemann, I. (2008). *Training of process control at the PCK refinery*. 11. Simulator training meeting, May 2008, Vienna, Austria.

Landeweerd, J. A. (1979). Internal representation of a process, fault diagnosis and fault correction. *Ergonomics, 22*, 1343–1351.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5*, 161–171.

Meyer, B. (2008). *The effects of computer-elicited structural and group knowledge on complex problem solving performance – An application of two computer-based tools for knowledge elicitation*. Open access doctoral dissertation, Humboldt University Press, Berlin.

Mohr, L. B. (1995). *Impact analysis for program evaluation* (2nd ed.). Thousand Oaks: SAGE.

Ormerod, T. C., Richardson, J., & Shepherd, A. (1998). Enhancing the usability of a task analysis method: A notation and environment for requirements specification. *Ergonomics, 41*, 1642–1663.

Rasmussen, J. (1990). Mental models and the control of action in complex environments. In D. Ackermann & M. J. Tauber (Eds.), *Mental models and human–computer interaction 1* (pp. 41–69). North-Holland: Elsevier.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment, 63*, 506–516.

Sauer, J., Burkolter, D., Kluge, A., Ritzmann, S., & Schüler, K. (2008). The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics, 51*, 953–967.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (2000). Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43*, 2043–2064.

Sauer, J., Wastell, D., & Hockey, G. R. J. (2000). A conceptual framework for designing micro-worlds for complex work domains: A case study of the Cabin Air Management System. *Computers in Human Behavior, 16*, 45–58.

Saus, E.-R., Johnson, B. H., Eid, J., Riisem, P. K., Andersen, R., & Thayer, J. F. (2006). The effect of brief situational awareness training in a police shooting simulator: An experimental study. *Military Psychology, 18*, 3–21.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67*, 605–610.

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

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27*, 285–300.

Shebilske, W. L., Goettl, B. P., & Garland, D. J. (2000). Situation awareness, automaticity and training. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 303–323). Mahwah: Lawrence Erlbaum Associates Publishers.

Shute, V. J., & Gawlick, L. A. (1995). Practice effects on skill acquisition, learning outcome, retention, and sensitivity to relearning. *Human Factors, 37*, 781–803.

Spiro, R. J., Feltovich, P. J., & Coulson, R. L. (1996). Two epistemic world-views: Prefigurative schemas and learning in complex domains. *Applied Cognitive Psychology, 10*, S51–S61.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 243–268). New York: John Wiley & Sons.

Vicente, K. J., Mumaw, R. J., & Roth, E. M. (2004). Operator monitoring in a complex dynamic work environment: A qualitative cognitive model based on field observations. *Theoretical Issues in Ergonomics Science, 5*, 359–384.

Vidulich, M. A. (2003). Mental workload and situation awareness: Essential concepts for aviation psychology practice. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 115–146). Mahwah: Lawrence Erlbaum Associates.

Wickens, C. D. (1998). Common sense statistics. *Ergonomics in Design*, *October*, 18–22.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton: CRC Press.

Wonderlic. (2002). *Wonderlic personnel test*. Libertyville: Wonderlic Inc.

Chapter 7

Study IV: Assessment of structural knowledge

Burkolter, D., Meyer, B., Kluge, A. & Sauer, J. (in press). Assessment of structural knowledge as a training outcome in process control environments. *Human Factors*.

Printed with permission by the Human Factors and Ergonomics Society.

Assessment of Structural Knowledge as a Training Outcome in Process Control Environments

Dina Burkolter, University of Duisburg-Essen, Germany, **Bertolt Meyer**, University of Zurich, Switzerland, **Annette Kluge,** University of Duisburg-Essen, Germany, and **Juergen Sauer**, University of Fribourg, Fribourg, Switzerland

Objective: The objective of the present studies was to apply a novel method for structural knowledge assessment to process control in order to assess the potential of its measures as a training outcome. **Background:** Traditionally, knowledge is assessed by verbal achievement tests on the subject matter. However, traditional methods are regarded as limited in their ability to assess higher order learning or understanding. **Method:** Two experiments (Experiment 1: N = 41; Experiment 2: N = 50) were conducted in which participants were given a 4-hour training session on a simulated process control task. At a later testing session, participants worked on the task for 70 minutes and completed knowledge tests on declarative, procedural and structural knowledge. Structural knowledge was measured using the computer-based association structure test (AST, Meyer, 2008), combining an association task and pathfinder network based on relatedness ratings. **Results:** In both studies, structural knowledge was significantly related to diagnostic performance, and evidence was found for internal consistency as well as convergent and predictive validity. **Conclusion:** Findings indicate that structural assessment with the AST shows promise as a training outcome in process control. **Application:** Potential applications of this research include the improvement of training design, delivery and evaluation.

Keywords: Process control systems, training, structural knowledge, knowl-edge elicitation

INTRODUCTION

Computer-based testing facilitates new measurements of knowledge, for example, the real-time rendering of items based on a participant's entries. The goal of the present study was to apply a novel test of structural knowledge to a process control environment in order to evaluate its potential as a training outcome.

Traditionally, knowledge is assessed by verbal achievement tests on the subject matter. A review by Goldsmith and Kraiger (1997) showed that the most popular methods to measure learning in working environments were paper-and-pencil tests. However, traditional methods are regarded as limited in their ability to assess higher order learning or understanding (Kraiger, Salas, & Cannon-Bowers, 1995). Stating that training outcomes are more complex and multifaceted, Kraiger, Ford and Salas (1993) proposed a theory-based classification scheme of learning outcomes in which the cognitive outcomes contain not only verbal knowledge such as declarative and procedural knowledge but also structural knowledge. Structural knowledge is the way in which individuals organize and interrelate concepts within a knowledge domain (Davis, Curtis, & Tschetter, 2003). It links declarative and procedural knowledge to mediate the transition in learning from 'knowing' to application (Hoole, 2006).

Initial studies assessed structural knowledge in the context of complex systems such as aviation or troubleshooting (e.g. Day, Arthur, & Gettman, 2001; Rowe, Cooke, Hall, & Halgren, 1996; Schvaneveldt et al., 1985) by employing the Pathfinder approach (Schvaneveldt, 1990) to structural assessment. Details on the Pathfinder approach and the associated measures are provided below.

Assessment of declarative, procedural and structural knowledge in the context of process control

The present study aims at extending research on structural knowledge to process control environments. Process control can be found in industries that regulate and control complex processes, such as chemical plants, nuclear power plants, or refineries. The processes generally involve a high number of interacting variables. Process control includes two major tasks: system control and stabilization on the one hand, and diagnosis and repair of fault states on the other (Wickens & Hollands, 2000). Successful process control performance depends on knowledge regarding procedures and how to operate the system, i.e. procedural knowledge, and on substantial knowledge of the system and its cause-effect relations, i.e. declarative knowledge (Kragt & Landeweerd, 1974; Wickens & Hollands, 2000). Kluwe (1997) measured declarative knowledge in process control using questions about plant components, their attributes and interactions ("What happens to the temperature in the mineral silo when...?", pp. 68/69) and found it to be related to system control performance. Furthermore, Kluwe's measurement of procedural knowledge, assessed by asking participants how to reach a certain goal, was correlated with declarative knowledge measures and system control performance. The measure required participants to mark sequences of inputs on a schematic representation of the interface.

While operating during normal states requires knowledge on 'what-leads-to-what-and-when', operating during fault states requires a more complex knowledge base involving the variety of ways the system could fail (Kluwe, 1997; Kragt & Landeweerd, 1974; Wickens & Hollands, 2000). For the latter, an operator has to draw on structural knowledge.

Structural knowledge was assessed, for instance, in the context of aviation for novice and expert fighter pilots using the Pathfinder approach (Schvaneveldt et al., 1985). Pathfinder delivers a graphical representation of a participant's knowledge structure by asking him or her to rate the similarity between domain-relevant concepts. Concepts are treated as graph nodes, and relatedness ratings of concepts are interpreted as an edge, with a distance label between the two. A high relatedness results in a short distance, and a low relatedness in a longer edge. The resulting graph is also referred to as PFNET. Schvaneveldt et al. showed that the PFNETs of novices and those of experts were distinguishable regarding their density – the number of edges divided by the number of possible edges in the PFNET. Novices were found to assume a higher number of links (i.e., a higher network density), while experts tended to identify the important, critical associations and thus exhibited a network with a lower number of links (i.e., a lower density).

Central to Pathfinder-based approaches to structural assessment is the assessment of the quality of the derived knowledge structure, for which the closeness measure (*C*) is most frequently employed (Davis et al., 2003). *C* is obtained by comparing a participant's PFNETs to a reference PFNET, and is based on the idea that the more similar the knowledge structure of a participant to that of an expert, the more likely this participant will display a similarly high performance (Goldsmith & Johnson, 1990).

The coherence measure of a PFNET is calculated without a referent structure (Schvaneveldt, Tucker, Castillo, & Bennett, 2001). Coherence is based on the assumption that if two concepts are rated as very similar, they must have similar relations to all other concepts. For each PFNET, the matrix of correlation estimates of edge values is correlated with the matrix of similarity ratings. The coherence therefore quantifies the extent to which a participant's ratings are coherent with the inherent logic of his PFNET. According to Schvaneveldt (2009), a coherence below .20 indicates too many inconsistencies. The coherence scores were able to differentiate the PFNETs of novices and experts (Schvaneveldt et al., 2001).

Despite the usefulness of PFNET-based structural assessment, two issues were raised by Meyer (2008): the demand for a fixed set of concepts that represent the knowledge domain, and the need for a reference network for *C*. As Pathfinder is based on fixed expert-identified concepts that are presented to all study participants, participants' familiarity with these concepts remains untested (Meyer, 2008). Furthermore, the reliance on expert-identified concepts presents a trade-off compared to free recall techniques, which allow a deeper understanding of a participant's knowledge (Davis et al., 2003).

Secondly, the *C* measure requires an expert network for reference. However, situations may arise in which such a network is unavailable, or too costly to elicit, and the validity of expert networks has not been addressed: Day et al. (2001) derived one referent structure based on the consensus of two experts and one based on averaged knowledge structures of two experts. There was a difference in the validity of the consensus judgments and average judgments with the latter predicting performance best. This finding underlines the validity issue of single expert networks as referent structures, since their validity would only be high if several experts' judgments were averaged. Accordingly, the present study applies a structural assessment technique combining a free recall technique with Pathfinder scaling operating on measures derived from the individual networks instead of comparisons with expert referent structures.

The Association Structure Test (AST)

The AST (Meyer, 2008) covers different facets of knowledge, primarily declarative and structural knowledge (see Table 1). It integrates an association task and pathfinder network scaling based on relatedness ratings into one IT-based test system. There are two parts, an *association task* and *relatedness ratings*.

TABLE 1. Overview of AST parameters

AST measure	Description	Data basis	Indicator for	Higher values indicate
Concepts	Number of associated concepts (nodes)	Association task	Verbalizable declarative knowledge	... more declarative knowledge
Clusters	Concepts following quickly after each other build a cluster	Association task (thinking times)	Differentiation of declarative knowledge	... more differentiated declarative knowledge
Edges	Number of connections between concepts	Rating task (and concepts from association task)	Verbalizable declarative knowledge with components of structural knowledge	... that the participant assumes more connections between concepts
Diameter	Longest path in the network	Rating task (and concepts from association task)	Span of knowledge	... wider spans of knowledge
Weighted density	Sum of all edge values divided by the number of all possible edges inside that network	Rating task (and concepts from association task)	Structural knowledge and structural implicitness (Dienes & Perner, 1999)	... that the network is more dense (and that the participant assumes more and/or stronger connections between concepts)

Association task Participants are asked to associate concepts that they think belong to a specified knowledge domain. In addition to recording the participants' associated concepts, thinking times during the word associations are logged. Drawing on the theory of spreading activation by Anderson (1983), semantically closely related terms are thought to follow quickly after each other, whereas semantically less related terms are assumed to result in longer pauses in thought between associations. Concepts form an interconnected network and retrieval is performed by spreading activation throughout the semantic network. Therefore, concepts are also referred to as nodes in the network. Activation of one concept will most likely activate an adjacent (i.e. semantically close) concept, while a long thinking time is associated with a longer propagation of the spreading activity through the network. Chase and Simon (1973) used the pauses in recall to identify chunks, based on the assumption that a pause would be related to the retrieval of a new chunk from memory. Longer pauses (more than 2 sec) were thereby associated with the retrieval of a new structure from memory, whereas a shorter pause (less than 2 sec) was associated with a succession of recalls drawn from the same structure. Beatty and Gerace (2002) asked physics students to associate to a topic area and recorded concept associations together with thinking times between associations. They found that clusters of concepts (i.e. concepts related to each other) were positively associated with exam scores, i.e. students whose associations were related performed better.

Data on thinking times of participants are employed by the AST to identify clusters of associated concepts: Concepts with a short thinking time between them form a cluster, while a long thinking time indicates a new cluster. The AST classifies the concepts into different clusters, with a cluster analysis over the thinking times between term associations. For each participant, long peaks of thinking times that stand out from short pauses between quick associations are identified based on average thinking time and its standard deviation for that particular participant. The detection of a peak is interpreted as separating two clusters of associated terms, i.e., the detection of a peak increases the number of clusters by 1 (Meyer, 2008).

Relatedness ratings The associated concepts are presented as pairs, and their relatedness is rated. The maximum number of concepts that are selected for pairwise comparisons can be determined in the AST's configuration. The total number of comparisons, $(n(n-1))/2$, depends on the number of previously associated terms (n). For example, if a participant entered 20 terms during the first stage of the test, he/she would have to perform 190 (20(20-1)/2) comparisons in the second stage. There is a tradeoff between reasonable comparison task length and its predictive validity. However, just ten concepts for pairwise comparisons have been shown to deliver an adequate prediction at a reasonable length of the comparison task (Davis et al., 2003). Therefore, a limit of 15 concepts is employed in the current studies.[1]

Participants do not have to label, describe or explain their judgment on the strength of the relatedness of two concepts. They rather make quick and intuitive decisions based on their gut-feeling, as one may appreciate a relationship between two concepts, but not necessarily carry a readily available label for this relation (Rothe & Warning, 1991). These relatedness ratings require neither a complex process nor a high degree of conscious processing. Therefore, the AST is thought to elicit relationships between knowledge elements that are difficult or impossible to verbalize, and thus capture a part of unconscious access to structural knowledge.

AST measures The AST converts a participant's matrix of concept relations into a PFNET ($n-1$, ∞) with the formulae supplied by Schvaneveldt (1990). As AST graphs are derived from word associations by the participants themselves, the graphs cannot be directly compared with each other or with referent graphs because these comparisons require graphs with the same nodes for all participants. Therefore, instead of a measure such as closeness (Goldsmith & Davenport, 1990), referent-free measures are employed to quantify AST-generated knowledge networks. Their value for performance prediction has been demonstrated by e.g. Bonato (1990) and Schvaneveldt et al. (2001).

The AST itself delivers a graph in an adjacency matrix form for each participant's PFNET. These graphs can be analyzed with a variety of graph analysis software packages to obtain graph-theoretic measures. In this study, we focus on five measures: the number of concepts (concepts), the number of clusters (clusters), the number of edges (edges), the diameter, and the weighted density. Extending Meyer's (2008) analyses, we also report Schvaneveldt et al.'s coherence measure (2001) and compare it to the AST measures.

The number of associated *concepts* that serve as graph nodes is seen as an indicator of declarative knowledge (i.e. the higher the number of concepts, the more declarative knowledge). According to Dienes and Perner's (1999) theory of implicit and explicit knowledge, the ability to verbalize a proposition indicates a certain level of explicitness. Thus, the number of concepts that a participant associates verbally to a stimulus indicates the magnitude of available declarative knowledge in the given subject domain. We report the number of associated concepts due to its close conceptual relation to declarative knowledge. As the participant is asked to associate concepts in the association task, the number of concepts is a face-valid indicator of participants' performance in the association task.

The *clusters* identified by the AST are thought to represent different sets of individual cognitive structures in the same knowledge domain – declarative knowledge (Meyer, 2008). A higher number of clusters indicates more differentiated declarative knowledge. The number of clusters is associated with the number of associated concepts. We report this as it has been shown to relate to performance (Beatty & Gerace, 2002).

A relatedness rating between two concepts is interpreted as an *edge* between the two concepts (i.e. edge value). Relatedness ratings are transformed into path distances for the Pathfinder scaling, resulting in short edges (i.e., 4 is recoded into 1) for strong relationships. Note that an edge is a set of two nodes. Edges cannot therefore be seen as independent of nodes, which in turn consist of the associated concepts. Thus, edges fit into the conceptualization of structural knowledge as something that links certain entities. The number of edges is seen as an indicator of what Dienes and Perner (1999) described as explicit knowledge, which "can be a representation of compounds (typically: compound properties) that leaves the structure of its components implicit" (p. 740). If the structure of the compound is implicit, we refer to it as structural implicitness in accordance with Dienes and Perner. A high number of edges indicates that the participant assumes more relations between concepts. We chose to report the number of edges in the context of this study because it is the prime unadjusted measure that carries structural information.

Note that despite the interconnection between nodes and edges and the resulting positive correlation between them, they are not mathematically deducible from each other: Participants specify nodes prior to specifying whether there is a connection between two nodes, which do not necessarily *have* to be connected. The number of edges inside a graph will thus correlate with the number of nodes, but it will not be deducible from the number of nodes.

To normalize the number of edges inside a graph (because the number of nodes can vary), the density measure is employed. This specifies the ratio of present edges in relation to possible edges inside the graph. The density is calculated by

$$\frac{l}{\frac{1}{2}n(n-1)}$$

where l is the number of lines (edges) in the graph and n is the number of nodes. The density measure will exhibit a negative correlation with the number of edges and nodes. The density measure approaches values close to 0 if very few edges are present in relation to the number of nodes, as Figure 1 illustrates.

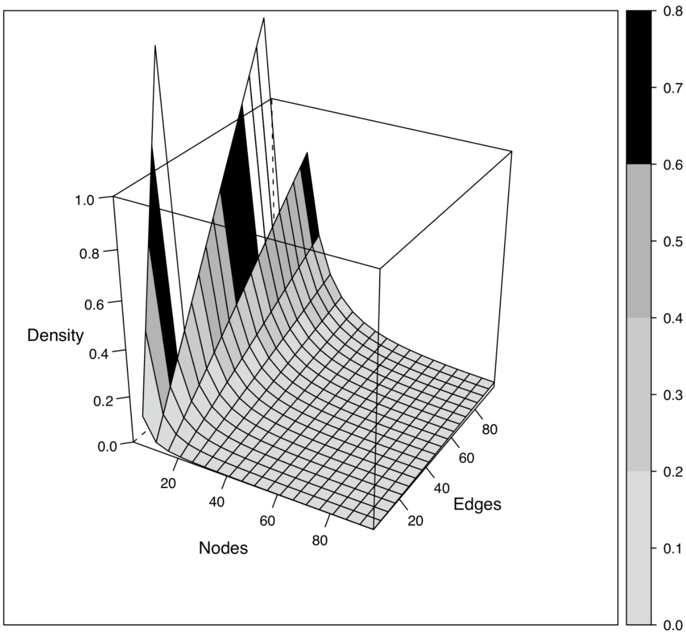

Figure 1. The density of a graph in relation to its number of nodes and edges. White areas are undefined for density values > 1. It becomes evident that the density measure approaches 0 for an increasing number of nodes and only visibly increases for a high number of edges at low numbers of nodes.

In order to increase the amount of structural information of the density measure, Meyer (2008) employed the *weighted density* measure, which takes the original edge values as delivered by the AST (higher values indicating a stronger connection) into account (as proposed by Benta, 2003). The weighted density is the sum of all edge *values* of the PFNET divided by the number of all possible edges inside that network, referring to the number of links between all nodes, without any pruning of the links by the Pathfinder algorithm. The weighted density of an undirected graph, such as a PFNET, is calculated by

$$\frac{2}{n(n-1)} \sum_{i=1;j=i+1}^{n} x_{ij}$$

where n is the number of nodes in the network and x_{ij} denotes the value of an undirected edge between nodes n_i and n_j. Contrary to the density, the weighted density is not deducible from the number of edges and nodes, as it takes a further argument – the edge values – into account. It will, however, also exhibit a negative correlation with the number of edges and nodes. In the case of the AST, where the strongest relation between two concepts is denoted with the value 4, the range of the weighted density is 0 to 4, with 4 indicating a graph where all possible connections are present and denoted with the highest connection strength. As the weighted density is the only proposed measure that relies on the number of edges in relation to the number of graph nodes *and* their strength inside a given graph, it carries the largest amount of structural information. We thus see it as the most appropriate indicator for structural knowledge, and, if applied to those edges that a participant cannot label, for structural implicitness.

The AST calculates the *diameter* of the generated PFNETs. The diameter is the longest path inside the graph delivered by Pathfinder scaling. Eckert (1998) treated the diameter as a measure for the span of structural knowledge, with higher values of the diameter indicating wider spans of knowledge. A graph with the same number of nodes (n) can have different diameter values up to $n - 1$. The diameter depends on the number of connections originating from nodes. If many nodes share many links to many other nodes, there will be short paths between a given node and any other given nodes, resulting in a shorter diameter of the graph. This would indicate that all nodes tend to be conceptually related to other nodes, i.e., that all nodes originate from similar knowledge domains. Finally, coherence (Schvaneveldt et al., 2001, see above) is compared to the AST measures.

A person with a large amount of declarative knowledge will associate a large number of concepts. Domain experts tend to indicate few but important links between concepts (Schvaneveldt et al., 1985). Furthermore, experts can experience difficulties in naming relations between concepts they assume to be present (Rothe & Warning, 1991). Therefore, skilled individuals are likely to place few but relevant connections – some implicit – between a potentially large number of concepts. As we assume that the weighted density measure can capture structural implicitness (see above) and will exhibit a negative correlation with the number of nodes in a graph (see Figure 1), a negative relationship between the weighted density measure and knowledge-based task performance can be expected. At the same time, we assume that experts know the few relevant connections that they place as well as the connections to other concepts, i.e. the network. Therefore, a low weighted density should co-occur with a high coherence, reflecting the awareness of the connections of all the concepts to each other.

In summary, the six graph-theoretic measures that we report were chosen because they either deliver direct operationalizations of the AST's two tasks, for theoretical reasons, i.e., the amount of structural information carried by the weighted density, and because they were employed successfully in the literature. Next, we will report studies addressing the validity of the presented claims before employing the AST in our experiments.

Studies on the validity of the association structure test

Internal consistency In order to address the internal consistency of the AST, we re-analyze a data set of 183 participants[2] who participated in various AST-related experiments obtained from Meyer (2008). The dataset consists of 102 male and 81 female students from a university and a vocational school in Germany (M = 24.6 years, SD = 3.3). As outlined above, all of the AST measures are assumed to correlate for theoretical and/or mathematical reasons. It is thus possible that the AST measures and coherence all measure the same underlying construct. As a single underlying construct would violate the assumption that the weighted density measure captures structural knowledge best, we assume a two-factor solution, with one factor representing structural knowledge and one factor representing verbalizable knowledge. We conducted a factor analysis with principal axis factoring and oblimin rotation (sigma = 0°) with all five AST measures and coherence across the sample. The Eigenvalue > 1 criterion, parallel analysis, optimal coordinate analysis and the scree plot all yielded a two-factor solution, which accounted for 57.2% of total variance (34.6% and 22.5%, respectively). Descriptive statistics, correlations and the rotated factor pattern matrix are reported in Table 2. As illustrated, participants' PFNETs had an average weighted density of 0.5. This means that on average, they placed about one weak edge (edge value = 1, prior to conversion to weights) for every two associated concepts. The average coherence of AST-PFNETs was 0.2 and therefore comparatively low, as Schvaneveldt (2009) assumed that coherence values below 0.2 indicate that the participant did not take the task seriously. However, his coherence values have to be interpreted in relation to edges that participants place between expert-identified concepts. In the AST, participants place edges between concepts that they themselves associated. Some of these links are assumed to be implicit (Rothe & Warning, 1991, see 1.2) and coherence might therefore be lower.

TABLE 2. Descriptive statistics, correlations, and pattern matrix of the rotated factor solution of the AST measures and coherence (oblimin rotation)

Variables	*M*	*SD*	(1)	(2)	(3)	(4)	(5)	Factor 1	Factor 2
1. Concepts (No.)	16.3	9.6	–					.90	–.09
2. Clusters (No.)	6.4	3.1	.76***	–				.76	–.08
3. Edges (No.)	36.4	24.3	.53***	.45***	–			.71	.26
4. Diameter (0 to [concepts –1])	2.6	0.8	.49***	.38***	.32***	–		.38	–.26
5. Weighted density (0 to 4)	0.5	0.5	–.52***	–.45***	–.08	–.44***	–	–.01	.99
6. Coherence (–1 to 1)	0.2	0.3	.45***	.36***	.16*	.32***	–.57***	.22	–.47

The number of nodes, edges, and clusters load substantially onto the first factor (all loadings > .70), whereas the weighted density is the only measure to exhibit a substantial loading on the second factor. Diameter and coherence do not show a substantial loading on any of the factors. This structure supports the rationale that the weighted density carries the largest amount of structural information, whereas the other measures are closely related to the amount of verbalizable knowledge, as indicated by the high loading of the number of concepts on the first factor.

In order to confirm this interpretation, we subjected the same data to a metric multidimensional scaling (Gower, 1966) with two dimensions, based on the Euclidean distances between AST measures, which were derived from the AST measures correlation matrix from the above sample (cf. Figure 2).

Coordinate 1 can be interpreted as a declarative – structural continuum, spanning between the number of nodes on the far left and the weighted density on the far right. The number of clusters is close to the number of associated concepts, because clusters can only be determined if a certain number of concepts are associated in the first place. The number of edges is placed further away from the nodes, because they incorporate a structural aspect, but are simultaneously dependent on the number of nodes (see above). The weighted density measure includes more structural information – the edge values – and is therefore placed further to the right.

Coordinate 2 can be interpreted as a quantitative (bottom) – qualitative (top) continuum. As discussed above, the diameter captures a more qualitative feature of a PFNET: the span or width of activated concepts. The coherence, placed at the top of the continuum, also denotes a quality of the PFNET: whether the links placed by the participant are consistent with the links one would expect, based on the correlations of edge values of linked concepts. The other AST measures, including the weighted density, are placed towards the bottom of the continuum, indicating a more quantitative character. The fact that the coherence measure is placed towards the left of the first continuum, which we labeled as declarative, can be interpreted insofar as consistent links are those edges that participants can explicitly label. As the weighted density of AST-generated PFNETs is assumed to capture structural implicitness to a certain extent (see above), it is placed further towards the right on the structural continuum.

In summary, both the factor analysis and the MDS support the theoretical rationale underlying the AST and especially the assumption that the weighted density can serve as an indicator of the number of structural links that participants assume to be present between concepts of which they are aware, in reference to the total number of concepts that comprise their domain knowledge. The negative correlation between coherence and the weighted density also supports the rationale behind the weighted density: If participants place few edges, these exhibit a high consistency in relation to the structure of the graph. At the same time, the two constructs are not the same, as demonstrated by the results of the correlation analysis, the factor analysis and the MDS.

Validity of measuring implicitness Meyer (2008) conducted further studies to assess the validity of the AST. A first experiment with a between-subjects design concerned the comparison of labeled and unlabeled relatedness ratings. In the labeled testing condition, participants could only indicate a relationship between concepts if they were able to explicitly label the kind of relation. If a value greater than zero had been chosen (zero indicating 'no/weak relationship'), a text box was displayed in which the nature of the relationship had to be explicated. In the unlabeled testing condition, participants were required to make quick and intuitive judgments on the strength of a relationship without describing the relationship. It was assumed that participants in the unlabeled testing condition would indicate a higher number of relationships (i.e. higher number of edges) than those in the labeled testing condition, as they could indicate relationships which they perceived without the need to have an explicitly available label for this relation (Rothe & Warning, 1991; see section 1.2). Thus, they could indicate all relations: those for which they could be able to provide an explicit label and those which they could not label. Participants in the labeled condition could only indicate those relationships for which they could provide an explicit label. Thirty undergraduate psychology students (19 female, M = 23.6 years, SD = 2.2) from a German university participated in the study. One group worked on the AST with labeled testing while the other group worked on the AST with unlabeled relatedness ratings, with the topic of a seminar as the stimulus concept. As expected, the number of edges was significantly higher in the unlabeled testing group (M = 53.5, SD = 45.4) than in the labeled testing group (M = 25.5, SD = 15.7; $t(24) = 2.25$, $p < .05$). With respect to the number of associated concepts, the labeled testing group (M = 10.9, SD = 5.7) did not differ significantly from the unlabeled testing group (M = 13.4, SD = 6.0; $t(24) = 1.2$, $p > .05$). Regarding the other measures of the AST, there were no significant differences between the two groups (all p's > .05). Hence, given a similar number of associated concepts, not having to verbalize an indicated relationship, resulted in a larger number of edges. These results indicate that quick and intuitive placement of relations between concepts might capture relations that a participant cannot verbalize and might be implicit.

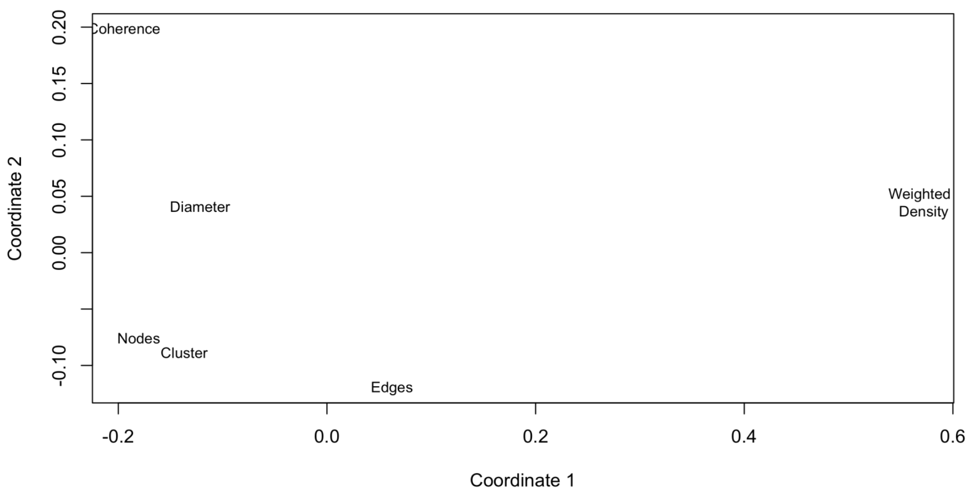

Figure 2. Metric multidimensional scaling of the five AST measures and coherence with two dimensions, based on the Euclidean distances between AST measures. Distances were derived from the AST measures' correlation matrix from an aggregated sample of participants (N = 183) who completed the AST in various contexts.

Predictive validity In an experiment to analyze predictive validity, fifty-two trainees (2 female, M = 20.3 years, SD = 3.5) from a vocational school for cleaners and plumbers in Germany were asked to complete the AST with professional terms. The vocational trainees worked on the AST several days after completing a written exam focusing on the stimulus concepts. Results show correlations between exam grade and concepts (r =.22, $p < .10$), clusters ($r = .25$, $p < .10$), edges ($r = .31$, $p < .05$) and diameter ($r = .18$, $p > .10$). Reanalyzing these data, we found no correlation between coherence and grade (r = .08, p > .10). The weighted density exhibited a negative correlation with participants' grades (r = -.36, $p < .05$): a less dense network was associated with a better grade. This finding concurs with the findings of Schvaneveldt and colleagues (1985), who found less dense Pathfinder networks among domain experts.

In summary, validation experiments as reported by Meyer (2008) show that the AST is able to tap into structural implicitness, and that the weighted density measure, which contains the most structural information, is related to knowledge-based performance. We therefore employ it for the elicitation of structural knowledge in the context of process control performance. In the first experiment, relationships between AST measures, other knowledge-related measures, and process control performance were determined in an exploratory way.

EXPERIMENT 1

Method

A data set extracted from a training experiment was used in the first study. The training experiment was conducted to study the effectiveness of three different training methods for process control performance. Findings concerning this research question are reported in a separate article (Burkolter, Kluge, Sauer, & Ritzmann, submitted). Data were collected two weeks after training.

Participants Forty-one students (four female) participated in the study. Participants were doing a B.Sc. in engineering at universities of applied sciences in Switzerland. They were paid 100 CHF (approx. 90 USD) for participation in all three sessions. The average age of the participants was 24.7 years (*SD* = 4.0).

The experimental task Process control performance was assessed through a computer-based simulation of a multi-task work environment (see Figure 2). The cabin air management system (CAMS; Sauer, Wastell, & Hockey, 2000) simulates a spacecraft's automated life support systems, but its underlying principles correspond to a process control task. CAMS consists of five main system variables (O_2, CO_2, cabin pressure, temperature, and humidity) that are maintained in normal range by automatic controllers. Two main tasks, system control and fault diagnosis, have to be accomplished in CAMS. The system control task requires the operator to intervene upon departure of a parameter from the target zone, either by adjusting the automatic control parameters or through manual control. Fault diagnosis involves the identification of the system disturbance by carrying out appropriate tests. The system fault can then be repaired by means of the maintenance facility. CAMS has already been employed in a range of different studies, and also in training experiments (e.g. Sauer, Burkolter, Kluge, Ritzmann, & Schüler 2008).

Testing procedure Knowledge and performance measures were all collected during the same testing session, which took place two weeks after the initial training. In this way, transfer of performance over a retention interval was measured, as transfer performance is an important factor in training research. The testing session on the CAMS task took 70 minutes, during which participants were to apply the acquired skills. Three fault states were included that were also part of the initial training and three novel fault states that were not addressed in the initial training. Thereafter, the written knowledge tests and the AST were conducted (approx. 45 minutes). Participants were given unlimited time to complete the written knowledge tests.

Measures The AST was employed with the stimulus word "CAMS". The maximum number of associated concepts to enter pairwise comparison was set to 15. Depending on the number of entered terms, the AST took about 20 to 30 minutes to complete. The graph analysis software UCINET (Borgatti, Everett, & Freeman, 2002), the igraph package (Csardi, 2009), and functions programmed in R (R Development Core Team, 2007) were used to obtain graph-theoretic metrics.

Declarative and procedural knowledge was assessed employing two previously employed and tested verbal knowledge tests on CAMS (see Sauer et al., 2008). Due to time constraints, a shortened version of the knowledge tests was conducted. The *declarative knowledge test* was composed of four multiple-choice items (e.g. "What happens to humidity when the heater is on?") with three alternatives ("increase", "decrease", "minimal or no effect"). The answer had to be explained in a subsequent open question. One item concerned the processes and relationships in CAMS ("Please explain which components or processes have an impact on cabin temperature and describe the mode of the relationship"). Answers to the free responses were compared to a solution that had been used in previous research (e.g. Sauer et al., 2008). Four components and processes (e.g. heating) had to be explained and each was credited with one point. The maximum possible score was eight points. The *procedural knowledge test* was developed following the rules of content-valid test design and assessed knowledge regarding fault descriptions, fault symptoms and repair steps based on the manual. The items referenced procedures of system control and fault repair, i.e. the steps needed to work through to control the system as well as to diagnose and repair system faults. In comparison to the declarative knowledge test in which understanding of the relationships between parameters and system components was required, the procedural knowledge test referred to the procedures that are relevant for accomplishment of the CAMS task. The test included two multiple-choice items concerning the description, two multiple-choice items concerning symptoms, and two multiple-choice items concerning interventions. A sample item regarding descriptions was: "Please state which fault is described: 'CO_2 scrubber operates with reduced effectiveness'" referencing to a specific procedure of fault repair. In items concerning fault symptoms, participants were required to state the system fault for a given symptom, while in items concerning fault repair, participants were required to state the corresponding system fault for a presented fault repair procedure. The maximum possible score was six points.

System control is one of the main tasks in CAMS. The five main system variables have to be maintained within normal range. If a parameter departs from this predefined zone, the operator needs to intervene. The duration in which parameters were in normal range was measured in seconds and converted into percentages.

The other main task in CAMS concerns *diagnostic performance*. Different system faults can be programmed into the simulation by the experimenter. The task of the operator is to diagnose and repair the system fault by means of the maintenance facility. The rate of correct diagnoses (diagnostic accuracy) is measured in percentages.

General mental ability (GMA) was assessed with the Wonderlic Personnel Test (Wonderlic Inc., 2002). The test comprises 50 items and captures verbal, numerical and spatial aspects of intelligence and learning aptitude. The participants had 12 minutes to work on the test which was conducted as a multiple-choice test.

Training The training sessions were typically carried out in small groups of four to six participants. All participants received a general introduction to CAMS through a multimedia-based instruction and were then given several minutes to explore the process control environment. Then, five system faults were introduced and, depending on the training approach, trainees did exercises as described in their instruction manual. The participants received either emphasis shift training (EST, $n = 13$), which was supplemented with a situation awareness training (EST/SA, $n = 14$), or drill and practice (D&P, $n = 14$). In the EST group, all participants were trained by alternately changing the emphasis on the two main tasks of process control (system control and diagnosis) from one exercise to the other (cf. Gopher, Weil, & Siegel, 1989). That is, participants practiced the actions of only one main task and did not have to execute the other interventions. The intended learning outcome of EST was the improvement of system control and diagnostic performance, especially for novel fault states and after a long retention interval. The EST/SA group received the same exercises as the EST group. Additionally, the EST/SA group was given an SA training session practiced through the freezing technique with debriefing (cf. Saus et al., 2006). This approach requires a simulated task to be randomly stopped and the trainee SA questions to be posed. EST/SA was designed to support especially diagnostic performance for novel fault states. The D&P group extensively practiced system control and diagnosis by repeating drills. Participants were required to follow the intervention steps closely. The intended goal of the D&P training was to enhance diagnostic performance of practiced fault states over a retention interval. Detailed information on the training approaches can be found in Burkolter et al. (2009).

A series of univariate ANOVAs were conducted in order to determine whether the different training sessions influenced performance in the written knowledge tests or the AST parameters. No significant differences between the training groups were found, either for the written knowledge tests (declarative: $F(2,40) = 1.685$, $p > .05$; procedural: $F(2,40) = 0.387$, $p > .05$) or for the AST measures (concepts: $F(2,40) = 1.215$, $p > .05$; clusters: $F(2,40) = 0.078$, $p > .05$; edges: $F(2,40) = 0.792$, $p > .05$; diameter: $F(2,40) = 0.667$, $p > .05$; weighted density: $F(2,40) = 0.226$, $p > .05$). The three training groups were thus combined into one sample ($N = 41$).

Results and discussion

Concepts that were most often associated included the five main parameters of CAMS, the context of CAMS (spacecraft, astronauts, life support system, simulation), the main tasks in CAMS (control, system faults, repair) and controllers (valves, graphs, time). The three most associated concepts (cabin pressure, O_2, control) were each associated by almost 50% of all participants, indicating an overlap of associated concepts.

Performance levels of system control and diagnostic performance (see Table 3) were comparable to results obtained in a previous, similar experiment with CAMS (Sauer et al., 2008).

All AST measures were intercorrelated in the manner assumed in section 1.3. Positive correlations between concepts, edges, clusters and diameter were found, whereas weighted density was negatively correlated with the other AST measures and coherence.

Surprisingly, the number of edges and the diameter exhibit negative correlations with declarative knowledge scores, and no AST measure correlates significantly with procedural knowledge scores. This finding could indicate validity issues of the employed test for declarative knowledge, as declarative knowledge scores show a near-zero correlation with process control performance. Therefore, the full versions of the written knowledge tests are employed in the second experiment.

To analyze convergent validity (cf. Kraiger & Jung, 1997), correlations between GMA , and the AST parameters were calculated. GMA was significantly and negatively related to the weighted density measure (i.e. the higher the mental ability, the less dense the knowledge network). This result applies to the finding that domain experts tend to place fewer relationships among concepts (Schvaneveldt et al., 1985; see 1.2), and experts will probably also exhibit higher GMA scores, as GMA and the acquisition of knowledge and expertise are related (Schmidt & Hunter, 1998).

System control performance was not correlated with any of the knowledge-related parameters. Diagnostic accuracy was significantly correlated with procedural knowledge scores and negatively related to weighted density. The negative correlation between diagnostic performance and the weighted density measure confirms previous findings (see 1.3; Schvaneveldt et al., 1985). There was also a negative, albeit not significant, correlation between the number of edges and performance. Thus, skilled individuals are assumed to exhibit a smaller density because they place only relevant (strong) connections (compare also Figure 3).

A hierarchical regression analysis for performance prediction was conducted (see Table 4). Procedural knowledge and coherence accounted for 15% of variance in diagnostic accuracy, and adding weighted density in a third step explained a further 6%. Even though this change in R^2 did not reach significance (p = .056), the results of the regression analysis indicate that weighted density explains variance in performance in addition to traditional knowledge measures.

TABLE 4. Summary of hierarchical regression analysis with diagnostic accuracy as a criterion

Variables	*B*	*SE B*	*β*	*p*
Step 1				
Procedural knowledge	4.82	2.01	.37	.021
Step 2				
Procedural knowledge	5.22	1.98	.40	.012
Coherence	14.63	8.89	.25	.109
Step 3				
Procedural knowledge	4.54	1.93	.35	.025
Coherence	3.22	10.31	.06	.757
Weighted density	–24.28	12.27	–.34	.056

Note. Adjusted R^2 = .11 for step 1; ΔR^2 = .04 for step 2; ΔR^2 = .06 for step 3.

EXPERIMENT 2

Based on the findings of the first experiment, for the second experiment we assume that (1) the pattern of results regarding the correlations and regression of the first study can be confirmed. Moreover, it is hypothesized that (2) procedural knowledge will show a positive relationship with diagnostic performance, (3) the weighted density of participants' AST-elicited PFNETs will correlate negatively with diagnostic performance, and (4) the weighted density of participants' AST-elicited PFNETs will explain incremental variance in diagnostic performance.

Method

The experimental task (CAMS) and the performance measures were the same as in Experiment 1 (see Section 2.1.2 for information on the experimental task and Section 2.1.4 for performance measures). The AST was also administered as in the first experiment (see 2.1.4).

Participants Fifty students (26 female) doing their B.Sc. and M.Sc. in engineering at a university in Germany participated in the study. The students were paid 100 EUR (approx. 130 USD) for participation in all three sessions. The mean age of the participants was 23.6 years (SD = 2.9).

Testing procedure Due to organizational and end-of-semester time constraints, the testing session could not be held two weeks after initial training, as in Experiment 1, but was held one week after training. Otherwise, testing was carried out in the same way as in Experiment 1.

Written knowledge tests As the validity of the shortened test was questioned in Experiment 1, the full version of the above-described knowledge tests (see 2.1.4) was administered. The *declarative knowledge test* comprised twelve multiple-choice questions requiring an additional short explanation and three open questions. The maximum possible score was 21 points. The *procedural knowledge test* contained twelve multiple-choice items (four items each concerning description, symptoms and interventions of fault states). The maximum possible score was twelve points.

Cognitive variables In addition to *GMA* (see 2.1.4), *working memory capacity* was assessed in Experiment 2, and was measured with a computerized n-back task. In the n-back task, the participant is requested to monitor a series of one-digit numbers from zero to nine presented in a random sequence. The participants are required to indicate whenever a number is presented that was presented two (2-back task) or three trials (3-back task) previously. Ten stimulus blocks with 24 stimulus trials each were administered, alternating between 2-back and 3-back conditions (only 2-back data were analyzed here; Schoofs, Preuss, & Wolf, 2008). Correct reactions are measured in percentages.

Training All participants received drill and practice as conducted in Experiment 1. The training took place in groups of usually ten participants with one supervisor. The same five system faults as in Experiment 1 were trained.

Results and discussion

The data of one participant were removed from the analyses. This participant named only two concepts in the association task, suggesting that he/she had not fully understood the objective of the AST.

Similarly to Experiment 2, the most associated concepts were also the parameters of CAMS, the context, the main tasks and controllers. The three most associated concepts (N_2, system fault, O_2) were each associated by two-thirds of all participants.

All AST measures and coherence were again intercorrelated (see Table 5) and showed the expected pattern (see 1.3).

As expected, GMA was again negatively correlated with weighted density. Working memory capacity was correlated with concepts and clusters. It seems reasonable that participants who are good at keeping information active while using it, i.e. had a high working memory capacity, were better able to associate concepts stored in long-term memory as chunks of similar and linked concepts (Wickens, Lee, Liu, & Becker, 2004).

Confirming results of the first experiment, system control performance was not related to any of the AST measures, but to procedural and declarative knowledge scores.

The second hypothesis was confirmed as procedural knowledge scores correlated with diagnostic accuracy. This result supports the assumption of validity issues with the shortened versions of verbal knowledge measures.

The third hypothesis was also supported by the results. Confirming the results of the first study, structural knowledge was significantly and negatively related to diagnostic accuracy. To provide a better understanding of the relationship between weighted density and diagnostic accuracy, the knowledge networks of a poor and a high performer regarding diagnostic performance are compared as examples. Figure 3 shows that the network of the poor performer contains fewer concepts than the network of the high performer, but with many edges (at least four per concept), while the network of the high performer was characterized by one main concept ('life support system'), which represents the main critical goal in CAMS, to monitor the crew's survival. This concept was linked to concepts critical for survival such as O_2, but not every concept had several connections to others.

Finally, a hierarchical regression analysis was performed, with diagnostic accuracy as a criterion and all knowledge measures that were correlated with it as predictors (see Table 6). Weighted density explained additional variance to declarative and procedural knowledge in diagnostic accuracy (23% in total). By adding coherence in the fourth step, even more variance in diagnostic accuracy was explained (30% in total, $p < .05$). However, the individual contribution of weighted density to the regression model decreased and was not significant ($p > .05$), while coherence displayed a higher individual contribution.

TABLE 3. Descriptive statistics and intercorrelations between variables of interest

Variables	*M*	*SD*	Min	Max	1	2	3	4	5	6	7	8	9	10
AST measures and coherence														
1. Concepts (No.)	12.8	8.2	4.0	45.0	–									
2. Clusters (No.)	5.5	2.7	2.0	16.0	.75**	–								
3. Edges (No.)	28.0	16.3	2.0	72.0	.36*	.33*	–							
4. Diameter (0 to [concepts –1])	2.4	0.8	1.0	4.0	.58**	.55**	.48**	–						
5. Weighted density (0 to 4)	0.5	0.3	0.0	1.5	–.56**	–.50**	–.07	–.60**	–					
6. Coherence (–1 to 1)	0.2	0.4	–0.9	0.9	.33*	.23	–.01	.38*	-.54**	–				
Cognitive variable														
7. GMA (0 to 50 pts.)	25.6	5.6	14.0	42.0	.19	.15	–.08	.15	–.31*	.22	–			
Written knowledge tests														
8. Declarative (% correct)	34.8	15.4	6.3	62.5	–.09	–.13	–.40**	–.33*	.08	–.06	.07	–		
9. Procedural (% correct)	66.3	27.3	0.0	100.0	–.10	.17	–.18	.11	–.09	–.12	.33*	–.08	–	
Process control performance														
10. System control (% correct)	73.0	9.4	47.6	83.6	.09	–.06	–.05	.03	.05	.13	.28^{+}	.01	.07	–
11. Diagnostic accuracy (% correct)	46.3	21.9	16.7	100.0	.11	–.05	–.20	.24	–.36*	.20	.47**	–.02	.33*	.25

Note. $^{+}$ = p < .10; * = p ≤ .05; ** = p < .01 (two-tailed). GMA: General mental ability.

TABLE 5. Descriptive statistics and intercorrelations between variables of interest

Variables	*M*	*SD*	Min	Max	1	2	3	4	5	6	7	8	9	10	11
AST measures and coherence															
1. Concepts (No.)	21.1	10.4	5.0	56.0	–										
2. Clusters (No.)	8.0	3.4	2.0	20.0	.79**	–									
3. Edges (No.)	33.1	17.3	4.0	98.0	.40*	.29*	–								
4. Diameter (0 to [concepts –1])	2.7	0.7	1.0	4.0	.41**	.38**	.21	–							
5. Weighted density (0 to 4)	0.3	0.3	0.0	1.7	–.57**	–.48**	.23	–.32*	–						
									–						
6. Coherence (–1 to 1)	0.3	0.3	–0.7	0.8	.41**	.29*	–.01	.17	.63**						
Cognitive and personality variables															
7. GMA (0 to 50 pts.)	25.8	6.3	10.0	40.0	.21	.28*	.07	–.05	–.35*	.38**	–				
8. Working memory (% correct)	81.8	15.3	35.4	99.0	.35*	.29*	.13	.17	–.22	.25^{+}	.27^{+}	–			
Written knowledge tests															
9. Declarative (% correct)	37.2	18.4	4.8	76.2	.08	.26^{+}	–.17	.02	–.18	.24^{+}	.43**	.16	–		
10. Procedural (% correct)	66.2	22.1	8.3	100.0	.06	.11	.00	.30*	–.06	.23	.19	.05	.43**	–	
Process control performance															
11. System control (% correct)	55.2	15.8	0.0	80.0	–.07	–.02	–.16	–.18	–.01	.07	.21	.14	.27^{+}	.38**	–
12. Diagnostic accuracy (% correct)	54.7	25.7	0.0	100.0	.07	.11	–.15	.02	–.28*	.47**	.26^{+}	.33*	.26^{+}	.47**	.58**

Note. $^{+}$ = p < .10; * = p ≤ .05; ** = p < .01 (two-tailed). GMA: General mental ability.

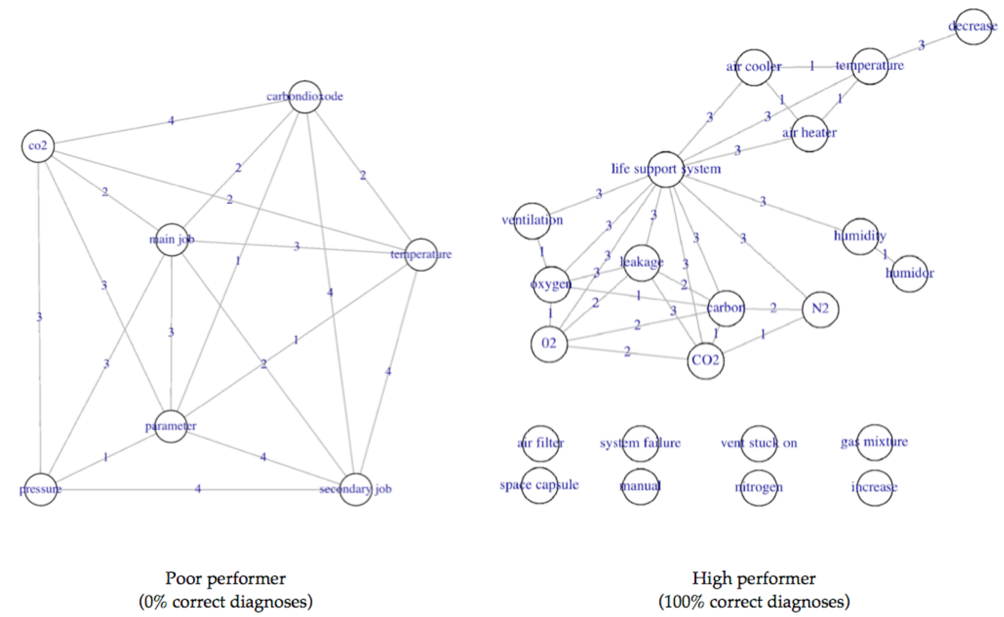

Figure 3. Knowledge elicited by the Association Structure Test graph for a poor and a high performer. 1 =strong relationship, 4 = no/weak relationship. Note that the high performer associated 22 concepts in the first part of the AST. Of these, 15 were presented for pairwise relatedness ratings. The participant did not rate one of the presented concepts as related to the others, resulting in a graph with 14 nodes. The concepts that were associated but not inserted into the graph are displayed at the bottom.

TABLE 6. Summary of hierarchical regression analysis with diagnostic accuracy as a criterion

Variable	*B*	*SE B*	*β*	*p*
Step 1				
Declarative knowledge	1.73	0.94	.26	.072
Step 2				
Declarative knowledge	0.49	0.96	.07	.616
Procedural knowledge	4.51	1.49	.44	.004
Step 3				
Declarative knowledge	0.17	0.95	.03	.855
Procedural knowledge	4.57	1.45	.44	.003
Weighted density	–19.68	10.38	–.24	.064
Step 4				
Declarative knowledge	0.03	0.91	.00	.978
Procedural knowledge	3.93	1.42	.38	.008
Weighted density	–1.53	12.72	–.02	.905
Coherence	28.62	12.55	.37	.027

Note. Adjusted R^2 = .05 for step 1; ΔR^2 = .14 for step 2; ΔR^2 = .04 for step 3; ΔR^2 = .07 for step 4.

GENERAL DISCUSSION

The aim of the present study was to provide a better understanding of knowledge use of operators in process control environments and to assess the potential of structural knowledge as a training outcome. The overall pattern of results in Experiment 2 largely confirmed findings of Experiment 1 and the hypotheses. Evidence for internal consistency as well as convergent, discriminant and predictive validity of the AST was obtained. First, both the factor analysis and the MDS supported the theoretic rationale underlying the AST and the assertion that weighted density carries the largest amount of structural information. Second, the AST measures were related to outside criteria such as coherence, GMA and working memory capacity. Third, the AST's weighted density measure was correlated with diagnostic performance.

Negative correlations between structural knowledge and diagnostic performance as well as the number of edges and performance were found in both experiments. Similar results were described by Meyer (2008), Schvaneveldt et al. (1985), and Rothe and Warning (1991). A comparison of the knowledge network depicted differences between a highly skilled participant and a low-skilled participant. Admittedly, there are differences in experience and expertise between domain experts, e.g. in the study by Schvaneveldt et al. and the participants in the present study. Therefore, there are limits on the transfer of results obtained with experts with several years of experience and the present study, which highlights the need for further research to include more experienced operators.

Diagnostic performance was not only correlated with procedural and declarative knowledge, but also with structural knowledge. System control, on the other hand, was only related to declarative and procedural knowledge. This finding indicates that different tasks of process control rely on different types of knowledge, corresponding to the assumption that "different control goals and states during process control call for quite different knowledge" (Kluwe, 1997, p. 62). While both system control and diagnostic performance are associated with the amount of knowledge acquired, only diagnostic performance seems to be based on the way knowledge is organized and concepts are related (cf. Day et al., 2001). This seems to relate to the fact that diagnosis requires a more complex knowledge base and thinking about what was caused by what, while controlling a system requires a focus on the forward flow of events (Wickens & Hollands, 2000; Landeweerd, 1979). Understanding the causes of a system fault might depend more strongly on implicit sections of knowledge and knowing different interrelations than predicting events. However, this assumption needs the further support of empirical evidence from studies measuring different types of knowledge in process control.

In the first experiment, weighted density added to explained variance in process control performance beyond the traditional measures of knowledge, a finding which is in line with previous research (e.g. Kraiger et al., 1995; Meyer, 2008). In the second experiment, these results were only partly confirmed. While coherence was not significantly related to performance in the re-analyzed study by Meyer (2008) and the first experiment, it was correlated with diagnostic performance in the second experiment and explained additional variance in performance. This pattern of results suggests the need for further research to compare weighted density and coherence and to assess in which cases which of the measures is appropriate. As the time between training and testing differed between Experiments 1 and 2, retention intervals might play a role in the usefulness of the two measures. However, the replicated finding that coherence and weighted density were related supports the notion that structural assessment is possible without expert-identified concepts and the necessity of a referent network, underlying the usefulness of the AST as a whole.

As the experiments did not include a pre-/post-test design with regard to structural knowledge, i.e. structural knowledge was not assessed before training but only after training, it is difficult to conclude that participants gathered more structural knowledge after training than before. However, as CAMS was new to all participants and has its own specific characteristics, which do not, or only partly, correspond to real physical processes, it would have been difficult for participants to associate any concepts to the stimulus 'CAMS' before training. Nevertheless, one could have employed a first testing of structural knowledge after an introduction to CAMS. We therefore see the cross-sectional design of Experiments 1 and 2 as a first step in validating the AST. Further studies should address this issue by employing a pre-/post-design.

The strengths of using participants' own associations for the relatedness ratings instead of presenting participants' preselected terms is that association tasks provide information about the organization and depth of a knowledge structure that cannot be captured by similarity judgments alone (Davis et al., 2003). However, users of the method might also have to deal with a trade-off of the procedure. As the core idea of the AST is to associate freely to a given stimulus term, a variety of what can be associated results, and not every term might seem appropriate. We assume, however, that the associated concepts are useful for the particular participant. Whether the subjective usefulness seen by the participants transfers to a general or "objective" usefulness is a question that, at least in our opinion, can be partly quantified by the weighted density: A detailed overall knowledge of the domain will lead to an association of more concepts and to fewer connections that the participant places with confidence, resulting in lower weighted density scores. Also, our analysis of the associated terms showed a reasonable overlap between participants, and demonstrated that terms were related to CAMS. This is a first indication that the participants associated reliable terms, but further research is needed.

With regard to training practice, the presented results indicate that structural knowledge as elicited by the AST might be useful in different ways. First, the assessment of structural knowledge allows for performance prediction regarding diagnostic performance and could be employed to differentiate between the levels of expertise among training participants. Second, the graphical representations of trainees' knowledge networks could be valuable for further training measures: Trainers could ask participants to explain certain connections between system components they think exist and possible misconceptions could be detected. Third, knowledge networks from subject matter experts might be used for visual instruction of novices and to allow for knowledge transfer – which is hardly feasible with traditional knowledge assessment methods. Furthermore, in a computer-based form of training in which measuring knowledge is also computer-based, there is no change of media, which might impede learning.

Altogether, considering the strengths and weaknesses of different knowledge assessment methods, the results of this study indicate that structural assessment with the AST shows promise as a training outcome in process control. However, further research, e.g. employing the AST in a field setting such as a plant, is needed for further evaluation.

ACKNOWLEDGMENTS

This project has been funded by the Swiss National Science Foundation (No. P001–106354).

REFERENCES

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Beatty, I. D., & Gerace, W. J. (2002). Probing physics students' conceptual knowledge structures through term association. *American Journal of Physics, 70*, 750–758.

Benta, M. I. (2003). *AGNA (Applied Graph & Network Analysis). Computer Software.* Department of Sociology, University College Cork, Ireland. Retrieved August 21, 2009, from http://www.geocities.com/imbenta/agna

Bonato, M. (1990). *Wissensstrukturierung mittels Struktur-Lege-Techniken: Eine graphentheoretische Analyse von Wissensnetzen* [Knowledge structuring with structure-laying-techniques: A graph-theoretical analysis of knowledge networks]. Frankfurt am Main: Peter Lang.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). UCINET: Software for social network analysis. Computer Software. Lexington, KY: Analytic Technologies.

Burkolter, D., Kluge, A., Sauer, J., & Ritzmann, S. (2009). *Comparative study of three training methods for enhancing process control performance: emphasis shift training, situation awareness training and drill and practice*. Manuscript submitted for publication.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Csardi, G. (2009). *Package 'igraph' – routines for simple graphs, network analysis. R library documentation*. Retrieved August 21, 2009, from http://igraph.sourceforge.net/doc/R/igraph
.pdf

Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: validity and utility of structural knowledge assessment. *Journal of Business and Psychology, 18,* 191–206.

Day, E., Arthur, W., & Gettman, D. (2001). Knowledge structures and the acquisition

of a complex skill. *Journal of Applied Psychology, 86*, 1022–1033.
Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 22, 735–808.
Eckert, A. (1998). *Kognition und Wissensdiagnose: die Entwicklung und empirische Überprüfung des computerunterstützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs-Technik (NET)* [Cognition and knowledge diagnosis: The development and empirical analysis of the computer-supported method for knowledge diagnosis Network-Elaboration-Technique (NET)]. Lengerich: Pabst.
Goldsmith, T.E. & Johnson, P.J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 241–254). Nowood, NJ: Ablex.
Goldsmith, T. E., & Kraiger, K. (1997). Applications of structural knowledge assessment to training evaluation. In J. K. Ford (Ed.), *Improving training effectiveness in work organizations* (pp. 73–96). Mahwah: Lawrence Erlbaum Associates.
Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 75–87). Nowood, NJ: Ablex.
Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica*, *71*, 147–177.
Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–328.
Hambrick, D.Z., Meinz, E., & Oswald, F.L. (2007). Individual differences in current events knowledge: Contributions of ability, personality, and interests. *Memory & Cognition, 35*, 304–316.
Hoole, E. (2006). *Integrating and evaluating mathematical models of assessing structural knowledge: Comparing associative networking methodologies*. Unpublished doctoral dissertation, Center for Assessment and Research Study, James Madison University, Harrisonburg, VA.
Kluwe, R. H. (1997). Acquisition of knowledge in the control of simulated technical system. *Le travail humain, 60*, 61–85.
Kragt, H., & Landeweerd, J. (1974). Mental skills in process control. In E. Edwards & F.P. Lees (Eds.), *The human operator in process control* (pp. 135–145). London: Taylor & Francis.
Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology Monograph, 78*, 311–328.
Kraiger, K., & Jung, K.M. (1997). Linking training objectives to evaluation criteria. In M. Quiñones and A. Ehrenstein (Eds.), *Training in a rapidly changing workplace: Applications of psychological research* (pp. 151–179). Washington: APA.
Kraiger, K., Salas, E., & Cannon-Bowers, J. A. (1995). Measuring knowledge organization as a method for learning during training. *Human Factors, 37*, 804–816.
Landeweerd, J. A. (1979). Internal representation of a process, fault diagnosis and fault correction. *Ergonomics, 22*, 1343–1351.
Meyer, B. (2008). *The effects of computer-elicited structural and group knowledge on*

complex problem solving performance. (Open access doctoral dissertation, Humboldt University Berlin, 2008). Retrieved August 21, 2009, from http://edoc.hu-berlin.de/docviews/abstract.php?id=29164

R Development Core Team (2007). *R: A language and environment for statistical computing*. Computer Software. Retrieved August, 21, from http://www.r-project.org.

Rothe, H.-J., & Warning, J. (1991). Zur Identifikation von semantischen Relationen im fachspezifischen Gedächtnisbesitz von Experten vermittels Analogiebildung [On the identification of semantic relations in experts' subject-specific memory by the means of analogy building]. *Zeitschrift für Psychologie, Supplement 11* , 375–384.

Rowe, A. L., Cooke, N. J., Hall, E. P., & Halgren, T. L. (1996). Toward an on-line knowledge assessment methodology: Building on the relationship between knowing and doing. *Journal of Experimental Psychology, 2*, 31–47.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment, 63*, 506–516.

Sauer, J., Burkolter, D., Kluge, A., Ritzmann, S., & Schüler, K. (2008). The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics, 51*, 953–967.

Sauer, J., Wastell, D. G., & Hockey, G. R. J. (2000). A conceptual framework for designing microworlds for complex work domains: A case study of the Cabin Air Management System. *Computers in Human Behavior, 16*, 45–58.

Saus, E.-R., Johnson, B.H., Eid, J., Riisem, P.K., Andersen, R., & Thayer, J.F. (2006). The effect of brief situational awareness training in a police shooting simulator: An experimental study. *Military Psychology*, *18*, 3–21.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schoofs, D., Preuss, D. & Wolf, O.T. (2008). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology, 33*, 643–653.

Schvaneveldt, R. W. (2009). *What is the coherence measure and when is it appropriate?* Retrieved August 21, 2009 from http://www.interlinkinc.net/FAQ.html#Coherence

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: studies in knowledge organization*. Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., et al. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23*, 699–728.

Schvaneveldt, R., Tucker, R., Castillo, A., & Bennett, W., Jr. (2001). *Knowledge acquisition in distributed mission training*. Paper presented at the 2001 Interservice/Industry Training, Simulation and Education Conference, Orlando, November.

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

Wickens, C.D., Lee, J.D., Liu, Y., & Gordon Becker, S.E. (2004). *An Introduction to*

Human Factors Engineering (2.ed.). Upper Saddle River: Pearson.
Wonderlic (2002). *Wonderlic Personnel Test*. Libertyville: Wonderlic Inc.

Footnotes

[1] If the number of terms entered during the first stage of the AST exceeds the specified maximum, the total number of terms selected for pairwise comparison in the second stage is equal to the specified maximum (i.e., 15). The sample of terms is chosen from clusters formed in the first stage: The very first term in each cluster enters the second stage; the remaining terms are selected randomly from each of the clusters in proportion to the cluster size. In this way, the selected terms represent the terms entered in the first stage and a preservation of the cognitive structure is maintained.

[2] The difference to the original sample size of 193 in Meyer's (2008) original study is due to missing values: Coherence is not properly defined for networks with three or less nodes, which occurred ten times.

Dina Burkolter is a research assistant for business and organizational psychology at the University of Duisburg-Essen, Germany. She completed her M.Sc. in Psychology at the University of Zurich, Switzerland in 2005.

Bertolt Meyer is a postdoctoral research fellow for social psychology at the University of Zürich, Switzerland. He obtained his Dr. rer.nat. in 2008 from Humboldt University in Berlin, Germany.

Annette Kluge is a Professor for business and organizational psychology at the University of Duisburg-Essen, Germany. She obtained her doctorate in ergonomics and vocational training at the University of Kassel, Germany in 1994.

Juergen Sauer is Associate Professor of cognitive ergonomics at the Department of Psychology, University of Fribourg, Switzerland. He obtained his Ph.D. from the University of Hull, United Kingdom, in 1997.

Chapter 8

Summary of results

In the next sections, the results of the four studies are summarized. First, results concerning effects of operator characteristics on training outcomes and second, results concerning effects of training methods on training outcomes will be described. Finally, results of all studies will be integrated and summarized on a comprehensive level.

8.1 Effects of operator characteristics on training outcomes

Studies I and II analyzed the effects of operator characteristics –both cognitive and personality variables– on training outcomes in process control. In Study I, cognitive ability (Wonderlic, 2002) and cognitive flexibility (Spiro, Feltovich & Coulson, 1996) were assessed as cognitive variables, and conscientiousness, openness, emotional stability (Saucier, 1994) and pretraining self-efficacy (Schyns & Collani, 2002) were assessed as personality variables. The objective of Study II was to confirm and extend findings of Study I by investigating the effects of the same and additional operator characteristics on training outcomes. Working memory capacity (Schoofs, Preuß & Wolf, 2008), set-shifting performance (Nagano-Saito et al., 2008; Nelson, 1976) and decision making (Brand et al., 2005) were assessed as additional cognitive variables, and need for cognition (Cacioppo & Petty, 1982; Keller, Bohner & Erb, 2000) and perfectionism (Altstötter-Gleich & Bergemann, 2006) were assessed as additional personality variables. For detailed information about the operator characteristics, their measurement and the design of the experiments, please refer to Studies I and II.

In Table 8.1[1] the effects of operator characteristics on training outcomes in process control are summarized based on the results of Studies I and II. The training outcomes include system control and fault finding performance as skill-based training outcomes as well as declarative and procedural knowledge as cognitive training outcomes.

Findings showed that cognitive ability was confirmed to be related to system control, fault finding and knowledge in more than one experiment. Moreover, further cognitive variables such as cognitive flexibility, working memory capacity, set-shifting performance and decision making were associated with process control performance. Regarding personality variables, need for cognition and perfectionism were related to knowledge. However, in none of the three experiments were the personality traits conscientiousness, openness, emotional stability or pretraining self-efficacy related to process control performance or knowledge.

8.2 Effects of training methods on training outcomes

In Studies III and IV, effects of training on training outcomes were investigated. While Study III analyzed the effect of different training methods on training outcomes, Study IV investigated how effects of training on training outcomes can be measured.

[1]The classification for cognitive and personality variables in Study I and Study II do not correspond fully. In Table 8.1, the classification from Study I was adopted.

Table 8.1: Evidence for positive effects of operator characteristics on process control performance (Study I and II)

	System control	Fault finding	Declarative knowledge	Procedural knowledge
Cognitive variables				
- Cognitive ability	+	++	++	+
- Cognitive flexibility	+	+		
- Working memory*		+		
- Set-shifting*[a]		+	+	
- Decision making*		+		+
Personality variables				
- Conscientiousness				
- Openness*				
- Emotional stability*				
- Self-efficacy*				
- Need for cognition*[b]			+	+
- Perfectionism*[c]			+	

Notes: + = evidence for positive relationship in one experiment; ++ = evidence for positive relationship in two experiments.*Only assessed in one experiment. [a]Ability to establish and then shift responses or tasks (Nagano-Saito et al., 2008). [b]Tendency to engage in and enjoy thinking (Cacioppo & Petty, 1982). [c]Subscale: Doubts about own actions. Procedural knowledge was not assessed in Study I.

In Study III, three training methods were designed to enhance attention management skills and process control performance. The objective was to support learners by providing them with attention management strategies in order to reduce their mental workload. The training methods EST, EST/SA and D&P were employed to examine performance in practiced and novel situations over a retention interval of six weeks.

The main results of the experiment are summarized as follows: First, D&P proved to be effective for fault finding of practiced faults in terms of speed and accuracy. Second, EST showed positive effects on system control performance during practiced faults. Finally, EST/SA supported system control performance during novel faults as well as diagnostic performance of novel faults.

Study IV aimed at applying a novel method for structural knowledge assessment to process control in order to assess its potential as a training outcome. In addition, the goal was to broaden the understanding of knowledge types needed for successful process control performance. Two experiments were conducted in which declarative, procedural and structural knowledge was assessed and process control performance was tested.

In both studies, structural knowledge was correlated with diagnostic performance. Diagnostic performance was not only correlated with procedural and declarative knowledge, but also with structural knowledge. System control, on the other hand, was only related to declarative and procedural knowledge. Evidence for internal consistency as well as convergent and predictive validity of the novel measure for structural knowledge was obtained.

8.3 Comprehensive summary of results of all studies

The research question formulated in the introduction was: How can training best support process control performance with regard to operator characteristics? And more specifically: Which operator characteristics, training methods and knowledge types support which training outcomes in process control best?

In Table 8.2, the main findings of all four studies are summarized according to the training objectives of process control training (see section 2.2). The Table depicts those operator characteristics, training methods and knowledge types that enhance system control and diagnostic performance best. The results of the four studies suggest the following:

– System control is best supported by cognitive ability and low cognitive flexibility, EST and EST/SA as well as enhancement of declarative and procedural knowledge.

– Fault finding is best supported by cognitive ability, high cognitive flexibility, high working memory capacity, high ability in set-shifting, EST/SA and D&P as well as enhancement of declarative, procedural and structural knowledge.

Table 8.2: Comprehensive findings of all studies: Enhancement of process control performance (indicated by a +) by operator characteristics, training methods and knowledge types

	System control	Fault finding
Operator characteristics		
Cognitive ability	+	+
Cognitive flexibility (CF)	+ (low CF)	+ (high CF)
Working memory		+
Set-shifting		+
Decision-making		+
Training methods		
Emphasis shift training (EST)	+ (practiced F.)	
EST and situation awareness	+ (novel F.)	+ (novel F.)
Drill and practice		+ (practiced F.)
Knowledge types		
Declarative	+	+
Procedural	+	+
Structural		+

Notes: F. = faults.

The findings suggest that the two main tasks of process control –system control and fault finding– are influenced in different ways by operator characteristics, training methods and knowledge types. A range of different operator characteristics affect performance, but do so differently regarding task components of process control. Working memory capacity, set-shifting performance and decision making affect fault finding as opposed to system control, which is not affected by these operator characteristics. Similarly, training methods affect the main tasks of process control differently. For example, EST supports system control performance but not fault finding.

Chapter 9

Discussion

9.1 Discussion of main findings

In accordance with previous findings (e.g. Colquitt et al., 2000; Salas et al., 2006b), operator characteristics were shown to be relevant for performance. Regarding GMA specifically, results were in line with findings by Schmidt and Hunter (1998) and Kramer (2009) found in different fields of work. The present results extend these findings on the relevance of GMA to the field of process control. Moreover, operator characteristics that are specifically related to process control performance were found. While Salas et al. (2006b) identified four individual characteristics, namely cognitive ability, self-efficacy, goal orientation and motivation, the presented studies also identified cognitive style, decision making and variables of executive functions (e.g. working memory capacity, set-shifting performance) as related to performance.

Taken together, cognitive operator characteristics as opposed to personality variables showed the greatest influence on process control performance. This finding is in line with a range of studies that emphasize the importance of cognitive skills, decision making and information processing in process control (Gatfield, 1999; Kragt & Landeweerd, 1974; Moray, 1997; Wickens & Hollands, 2000). The results also correspond to findings obtained by

Ackerman, Kanfer and Goff (1995) in the context of air traffic control. Ackerman et al. investigated the skill acquisition of over ninety trainees in an air traffic controller simulation task (Terminal Radar Approach Controller, TRACON). While Ackerman et al. found significant relationships between cognitive ability and performance in TRACON, there was no significant relationship between personality traits ("Big Five") and performance. Hence, similarly to the present findings, cognitive variables showed a stronger influence on performance than personality variables on performance.

The cognitive operator characteristics cognitive ability, cognitive flexibility, working memory capacity, set-shifting performance and decision making showed an influence on performance and knowledge acquisition. These cognitive operator characteristics can be related to the process control task and its requirements in order to gain a better understanding of their mode of action and effectiveness. Considering the task and demands of process control, cognitive operator characteristics can be related to different stages and functions in the control of the system. Cognitive ability can help in acquiring a general understanding of the system and the plant. Cognitive ability supports (a) the acquisition of knowledge and the development of a mental model of the plant (cf. Schmidt & Hunter, 1998), supports (b) the understanding of the causal structure of the system, and (c) the understanding of instructions and information given by the system (Kluge, 2008; Wittmann & Hattrup, 2004). Depending on the level of cognitive flexibility, either system control or fault finding performance are mostly supported. High working memory capacity can aid fault finding, because information about the system state, symptoms and their effects can be memorized better, so that information can be analyzed, steps can be planned and actions taken (cf. Burkolter et al., 2007; Ormerod, Richardson & Shepherd, 1998). Similarly, high set-shifting ability can help in identifying patterns of symptoms and responding to them in order to diagnose a system fault. Moreover, with high set-shifting ability, an operator is also able to flexibly shift attention to another task or symptom when needed (cf. Nagano-Saito et al., 2008). In this way, set-shifting ability may also help in fault finding of novel system faults, i.e. adaptive transfer. Finally, high dynamics, interrelatedness, feedback delays, and opaqueness pose high demands on decision making in complex task performance (Gonzalez, 2004; Gonzalez et al., 2005). Therefore, high decision making ability is helpful in the interaction with complex and dynamic systems (e.g. Wickens et al., 1998) and it seems reasonable to assume low risk-taking behavior to be more effective in process control than high risk-taking behavior.

Results regarding the effects of operator characteristics on performance are important not only for personnel selection and training, but also for interface design. As Szalma (2009) notes, in the future, it will also be important to meet the needs of the individual in interface design. By incorporating an individual differences approach to human factors research and practice, the understanding of human–machine interaction can be improved and the accuracy of theoretical models of human and system performance can be increased. This knowledge can be used to improve interface design by making the designs more inclusive (Szalma, 2009). For instance, one could think of different configurations of displays with respect to the cognitive styles of operators depicting low or high numbers of parameters in different layouts.

Regarding training methods in process control, D&P was effective for speed and accuracy in diagnosing practiced fault states. D&P has much in common with procedure-based training in that both training approaches emphasize the necessity to work through the steps of a procedure precisely (Sauer et al., 2000b; Hockey et al., 2007). Basically, the two approaches differ in the number of exercises given to trainees with D&P focusing on a high number of drills and repetition. In line with findings on the effectiveness of procedure-based training, the effectiveness of D&P was confined to familiar fault states. Hockey et al. state that training approaches concentrating on teaching procedures rather than teaching knowledge have a restricted range of transfer. In contrast, EST/SA was successful both at supporting system control performance during novel faults and at fault finding of novel faults. SA training facilitates the construction of mental models (Endsley & Robertson, 2000) which might have supported the transfer from practiced faults to novel faults. Similarly, knowledge-based training enhances the construction of mental models, which is assumed to be more flexible for applying learning to unfamiliar situations (Hockey et al., 2007).

The findings of Study IV showed that structural assessment (using pathfinder analyses) was predictive for performance and transfer. These results are in line with findings from studies entailing complex and dynamic tasks such as aviation or troubleshooting (e.g. Day, Arthur & Gettman, 2001; Rowe, Cooke, Hall & Halgren, 1996). These research findings are thus extended by similar results from process control. Despite a range of studies concerning structural knowledge and structural assessment in educational psychology (e.g. Goldsmith, Johnson & Acton, 1991) which have also influenced research in organizational psychology (e.g. Davis et al., 2003), there has been much less research in the field of process control on structural knowledge and its assessment. Studies on knowledge and knowledge types related to process control are generally harder to come by (Kluwe, 1997). However, as not only studies in the field of complex and dynamic tasks, but also Study IV, showed structural assessment as a training outcome to be a promising avenue, pursuing this research should be encouraged.

9.2 Discussion of results found in all studies

In the following sections, results found in all studies are discussed with regard to previous findings, limitations of the studies, and further research and practice.

9.2.1 The importance of designing training according to the main tasks of process control

First of all, the most important finding is that Landeweerd's (1979) notion that abilities in system control and fault finding are independent was confirmed in all of the studies. In the study on operator characteristics, system control and diagnostic performance were supported by different levels of cognitive flexibility. Further studies on operator characteristics showed that only fault finding was related to working memory capacity, set-shifting performance and decision making while system control performance was not associated with these characteristics. Moreover, the comparative study of training methods revealed that the main tasks of process control were influenced differently by training. Emphasis shift training, for instance, showed only effects on system control performance (during practiced fault states). Finally, system control and fault finding seem to depend on different knowledge types. While both tasks were associated with declarative and procedural knowledge, only fault finding was related to structural knowledge.

These findings first of all emphasize the importance of clearly stating system control and fault finding performance separately as a training objective in process control, since they are supported by different operator characteristics, training methods and knowledge types. Second, training interventions need to be designed according to each of the subtasks.

9.2.2 The importance of assessing training outcomes in a multifaceted way

The findings of the studies yield a differentiated and multifaceted picture of factors that influence process control performance on an individual level. All studies showed that the effects of the independent variables on the training outcomes varied depending on the type of training outcome. This applied not only to skill-based outcomes as described above (i.e. system control and fault finding), but also for cognitive and affective outcomes.

Studies I and II showed that operator characteristics had different effects on system control and fault finding. In addition, operator characteristics affected declarative and procedural knowledge differently. For instance, some operator characteristics had only an effect on declarative knowledge but not on procedural knowledge and vice versa (e.g. set-shifting performance, decision making). In Study III, no effects of training methods on cognitive and affective outcomes were found; effects were found only on skill-based training outcomes. Finally, Study IV clearly showed that assessing structural knowledge as a further cognitive training outcome explained additional variance in process control performance. Hence, it seems promising to include additional measures of training outcome in training evaluation.

In conclusion, the findings of this thesis show that training outcomes have to be assessed in a multi-faceted way. In this way, it can be ensured that the effects of training are evaluated in a comprehensive and differentiated manner. This finding supports suggestions by Kraiger et al. (1993), who stated that learning outcomes are multidimensional. Therefore, they see it as unnecessarily restrictive and out of step with modern learning theories to solely measure changes in verbal knowledge or behavioral capacities as training outcomes. The present studies have all shown, on an empirical basis, that the latter also holds true for process control performance.

9.2.3 The importance of the training analysis phase in training development

Taken together, the two main conclusions drawn from the results of the studies above, highlight the importance of the training analysis and design phase. First, the importance of designing training according to the main tasks of process control and second, the importance of assessing training outcomes in a multifaceted way was derived from the results. These two conclusions give rise to a third conclusion: The importance of the training analysis phase in training development.

Findings indicate that process control performance can be supported by a thorough training analysis that takes into account the differentiated picture of factors influencing different training outcomes. To be able to (a) design training and (b) plan the measurement of training outcomes appropriately, the task characteristics and its requirements have to be determined thoroughly. Individual characteristics, characteristics of the task and its requirements as well as a multidimensional measurement of training outcomes have to be considered beforehand. With knowledge gained in a training and task analysis, effective training and a multi-faceted training evaluation can be designed.

Goldstein (1993), Patrick (1992) and Salas et al. (2006b) emphasized the importance of a thorough training needs analysis in training development. The present studies support this recommendation based on empirical data drawn from several training experiments. Moreover, this recommendation for training design in general but not specified for a work environment, also seems to be relevant for the field of process control, i.e. a complex and potentially hazardous work environment.

9.3 Limitations of the studies

All studies of this thesis were conducted with the computer-based process control task CAMS. There are several reasons in favor of using CAMS as an experimental task, such as the possibility to measure adaptive transfer and primary and secondary task performance. On the other hand, the use of simulations in general and CAMS in particular also entails some trade-offs. Taking into account that CAMS has been applied in a range of studies, it might seem surprising that there has not been any formal validation of the simulation (cf. Sauer et al., 2008). A validation study, for example comparing the CAMS task to another process control task in order to analyze discriminant and convergent validity, might have been beneficial (cf. Burkolter, Kluge, German & Grauel, 2009). As the use of simulations involves the question of generalizability to the "real" world, a sound validation of a simulated environment supports external validity. However, an important requirement for generalization from simulation studies is theory (Brehmer, 2004). This requirement is given as CAMS was developed on the basis of a theory of human performance (Sauer et al., 2000a).

Common to all studies was that either trainee operators or engineering students were invited as participants. Thus, individuals at the beginning of their vocational careers, who so far have little experience, participated in the studies. Presumably, effects of operator characteristics and training on performance might be different with more experienced operators (cf. Study II), especially if one considers that in process control, operators often work for several years on a plant and gather a great deal of experience before they are given their assignments as control room operators (Kluge et al., 2009). Thus, the transfer of results to process control is limited to novices and initial learning. Moreover, participants did not practice the task between the experimental sessions, which might not directly apply to real-world settings, where operators usually work between training sessions and thus gain experience in this time.

9.4 Suggestions for further research and practice

The independence of skills in system control and fault finding was replicated in all experiments. This is crucial information for training. One needs to consider the independence of the two tasks and the findings of the studies in different phases of training design and delivery. Following the training phases by Salas et al. (2006b), the following suggestions are made:

– In the *training analysis phase*, training objectives and cognitive requirements should be elaborated separately for fault finding and system control.

– In the *training design phase*, the focus should be on operator characteristics that influence learning outcomes in process control tasks differently. However, the trainability of operator characteristics is limited. Nevertheless, insights into essential operator characteristics can guide decisions in selection – for example, in order to answer the question: Who will be most likely to succeed in a training program?

– In *training development and delivery*, training should contain parts especially designed for enhancing fault finding skills and system control performance. For fault finding of practiced fault states D&P and for fault finding of novel fault states, EST combined with situation awareness training is suggested. System control during practiced faults can be supported with EST, while EST combined with situation awareness training supports system control during novel faults.

– In *training evaluation*, it will be important to cover different facets of learning and knowledge as training outcomes, as they are related differently to the main tasks of process control.

Regarding future research, an important fact to consider are long-term effects – as much for operator characteristics as for training and training evaluation. Since there are situations in process control that may not occur for several years (see section 2.2), long-term effects are of special importance in this field. Thus, future experiments might explore the predictive qualities of operator characteristics for process control performance over several months or years. Likewise, the effects of different training methods might be evaluated after a retention interval of a number of months or years.

Regarding operator characteristics, not only might the direct influence of a variable on process control performance be of interest; the interaction of two operator characteristics on performance, i.e. the contribution of two variables at the same time, might explain additional variance in performance. Moreover, participants might benefit differently from training with respect to operator characteristics and different training methods (e.g. Gully, Payne, Koles & Whiteman, 2002). Hence, further research might investigate interaction effects between training methods and individual differences (aptitude–treatment interaction effects; e.g. Goldstein, 1993; Patrick, 1992).

Furthermore, the effects of different training methods on process control performance might be analyzed with respect to their sequencing in an overall training program, i.e. to analyze which training method may best support skill and knowledge acquisition in the beginning or at a later stage of skill acquisition. This suggestion is also related to the question of how the presented results were influenced by the fact that the study participants were novices. Future research could explore which training methods are best suited for which skill acquisition phase.

In conclusion, this doctoral thesis has shown that the integration of research from different disciplines is fruitful and can lead to new insights. It is worthwhile to reinforce the study of individual differences in human factors research. Given a changing workforce and increased diversity at the workplace, for instance regarding age or ethnic background (Hedge & Borman, 2006; Thayer, 1997), a consideration of individual differences becomes more important in training research. Furthermore, it is assumed that the use of technology will spread in the future, leading to jobs that require fewer sensory and physical skills and more (complex) cognitive skills (Hedge & Borman, 2006). This development also has implications for training design and research. By integrating findings from the training of complex, cognitive skills into complex and dynamic work domains, present training approaches can be enhanced in order to meet future demands.

References

Ackerman, P., Kanfer, R. & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, *1*, 270–304.

Aguinis, H. & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, *60*, 451–474.

Altstötter-Gleich, C. & Bergemann, N. (2006). Testgüte einer deutschsprachigen Version der Mehrdimensionalen Perfektionismus Skala von Frost, Marten, Lahart und Rosenblate (MPS-F). [Psychometric properties of a German version of the multi-dimensional perfectionism scale by Frost, Marten, Lahart and Rosenblate (MPS-F)]. *Diagnostica*, *52*, 105–118.

ASTD. (2009). *2009 State of the industry report.* Alexandria: American Society for Training and Development.

Badke-Schaub, P., Hofinger, G. & Lauche, K. (2008). *Human Factors. Psychologie sicheren Handelns in Risikobranchen [Human factors. The psychology of safe behavior in risk organizations].* Heidelberg: Springer.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*, 775–779.

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J. & Markowitsch, H. (2005). Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: associations with executive functions. *Neuropsychology*, *19*, 267–277.

Brehmer, B. (2004). Some reflections on microworld research. In S. Schiflett, L. Elliott, E. Salas & M. Coovert (Eds.), *Scaled worlds: Development, validation and applications* (p. 22–36). Aldershot: Ashgate.

Burkolter, D., Kluge, A., German, S. & Grauel, B. (2009). Waste Water Treatment Simulation (WaTr Sim): Validation of a new process control simulation tool for experimental training research. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (p. 1969–1973). Santa Monica: HFES.

Burkolter, D., Kluge, A., Schüler, K., Sauer, J. & Ritzmann, S. (2007). Cognitive requirements analysis to derive training models for controlling complex systems. In D. de Waard, G. Hockey, P. Nickel & K. Brookhuis (Eds.), *Human factors issues in complex system performance* (p. 475–484). Maastricht: Shaker Publishing.

Cacioppo, J. & Petty, R. (1982). The need for cognition. *Personality and Social Psychology*, *42*, 116–131.

Cannon-Bowers, J., Rhodenizer, L., Salas, E. & Bowers, C. (1998). A framework for understanding prepractice conditions and their impact on learning. *Personnel Psychology*, *51*, 291–320.

Chmiel, N. (2000). *Introduction to work and organizational psychology: A European perspective.* Oxford: Blackwell Publishers.

Colquitt, J., LePine, J. & Noe, R. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85(5)*, 678–707.

Davis, M., Curtis, M. & Tschetter, J. (2003). Evaluating cognitive training outcomes: validity and utility of structural knowledge assessment. *Journal of Business and Psychology*, *18*, 191–206.

Day, E., Arthur, W. & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, *86*, 1022–1033.

Dipboye, R. (1997). Organizational barriers to implementing a rational model of training. In M. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace. Applications of psychological research* (pp. 31–60). Washington, D.D.: American Psychological Association.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64.

Endsley, M. & Robertson, M. (2000). Training for situation awareness in individuals and teams. In M. Endsley & M. Robertson (Eds.), *Situation awareness analysis and measurement* (p. 349–365). Mahwah: Lawrence Erlbaum Associates Publishers.

Fahlbruch, B., Schöbel, M. & Domeinski, J. (2008). Sicherheit [Safety]. In P. Badke-Schaub, G. Hofinger & K. Lauche (Eds.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen* (p. 19–35). Heidelberg: Springer.

Fincham, R. (1999). *The individual, work, and organization.* Oxford: Oxford University Press.

Flexman, R. & Stark, E. (1987). Training simulators. In R. Flexman & E. Stark (Eds.), *Handbook of Human Factors* (pp. 1012–1038). New York: Wiley.

Flin, R., O'Connor, P. & Mearns, K. (2002). Crew resource management: improving team work in high reliability industries. *Team Performance Management*, *8*, 68–78.

Ford, J. (1997). Advances in training research and practice: An historical perspective. In J. Ford (Ed.), *Improving training effectiveness in work organizations* (p. 1–18). Lawrence Earlbaum Associates: Mahwah.

Gatfield, D. (1999). Can cognitive science improve the training of industrial process operators? *Journal of Safety Research*, *30*, 133-142.

Gillmann, W. (2009, February 25). Die Renaissance der Atomkraft. [The renaissance of nuclear energy]. *Handelsblatt*.

Goldsmith, T., Johnson, P. & Acton, W. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, *83*, 88–96.

Goldstein, I. (1993). *Training in organizations.* Pacific Grove: Brooks/Cole Publishing Company.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*, 449–460.

Gonzalez, C., Vanyukov, P. & Martin, M. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*, 273–286.

Gopher, D. (1996). Attention control: explorations of the work of an executive controller. *Cognitive Brain Research*, *5*, 23–28.

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

Gully, S., Payne, S., Koles, K. & Whiteman, J. K. (2002). The impact of error training and individual differences on training outcomes: An attribute–treatment interaction perspective. *Journal of Applied Psychology*, *87(1)*, 143–155.

Hedge, J. & Borman, W. (2006). Personnel Selection. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 458–471). New York: John Wiley & Sons.

Hockey, G. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*, 73–93.

Hockey, G., Sauer, J. & Wastell, D. (2007). Adaptability of training in simulated process control: Knowledge versus rule-based guidance under task changes and environmental stress. *Human Factors*, *49*, 158–174.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method.* Oxford: Elsevier Science Ltd.

IAEA. (2004). *Use of control room simulators for training of nuclear power plant personnel.* Vienna: IAEA.

IAEA. (2009). *Nuclear safety review for the year 2008.* Vienna: IAEA.

Keller, J., Bohner, G. & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung - verschiedene Prozesse? Präsentation einer deutschen Fassung des "Rational-Experiential Inventory" sowie neuer Selbstberichtskalen zur Heuristiknutzung. [Intuitive and heuristic decision making - different processes? Presentation of a German scale of the 'Rational Experiential Inventory' and new self-report scales of using heuristics.]. *Zeitschrift für Sozialpsychologie*, *31*, 87–101.

Kirckpatrick, D. (2007). *Evaluating training programs (3rd ed.).* San Francisco: Barrett-Koehler.

Kluge, A. (2008). What you train is what you get? Task requirements and training methods in complex problem-solving. *Computers in Human Behavior*, *24*, 284–308.

Kluge, A., Sauer, J., Schüler, K. & Burkolter, D. (2009). Designing training for process control simulators: A review of empirical findings and current practices. *Theoretical Issues in Ergonomics Science*, *10*, 489–509.

Kluwe, R. (1997). Acquisition of knowledge in the control of simulated technical system. *Le travail humain*, *60*, 61–85.

Kragt, H. & Landeweerd, J. (1974). Mental skills in process control. In H. Kragt & J. Landeweerd (Eds.), *The human operator in process control* (pp. 135–145). London: Taylor & Francis.

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

Kraiger, K., Salas, E. & Cannon-Bowers, J. (1995). Measuring knowledge organization as a method for learning during training. *Human Factors*, *37*, 804–816.

Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland. Vertiefende und weiterführende Analysen. [General intelligence and professional success in Germany. Extended and continued analyses.]. *Psychologische Rundschau*, *60*, 82–98.

Landeweerd, J. (1979). Internal representation of a process, fault diagnosis and fault correction. *Ergonomics*, *22*, 1343–1351.

Lane, S. (2007). A historical view of human factors in the united states. In L. Koppes (Ed.), *Historical perspectives in industrial and organizational psychology* (p. 243–263). Mahwah: Lawrence Erlbaum Associates.

Mannarelli, T., Roberts, K. & Bea, R. (1996). Learning how organizations mitigate risk. *Journal of Contingencies & Crisis Management*, *4*, 83–92.

Manzey, D. (2008). Systemgestaltung und Automatisierung [System design and automation]. In P. Badke-Schaub, G. Hofinger & K. Lauche (Eds.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen* (p. 307–324). Heidelberg: Springer.

Meshkati, N. (1991). Human factors in large-scale technological systems' accidents: Three Mile Island, Bhopal, Chernobyl. *Industrial Crisis Quarterly*, *5*, 131–154.

Meyer, B. (2008). *The effects of computer-elicited structural and group knowledge on complex problem solving performance. Open access doctoral dissertation, Humboldt University Berlin.* Retrieved September 4, from http://edoc.hu-berlin.de/docviews/abstract.php?id=29164.

Meyer, C., Nachtwei, J. & Kain, S. (2009). Der moderierende Effekt von Persönlichkeitsmerkmalen auf die Leistung von Operateuren bei der Prozessführung und -überwachung in komplexen Mensch-Maschine-Systemen [The moderating effect of personality characteristics on performance of process control operators in complex man-machine systems]. In *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme* (p. 185–190). Düsseldorf: VDI Verlag.

Moray, N. (1997). Human factors in process control. In N. Moray (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1944–1971). New York: John Wiley & Sons.

Morris, N. & Rouse, W. (1985). Review and evaluation of empirical research in troubleshooting. *Human factors*, *27*, 503–530.

Muchinsky, P. (1987). *Psychology applied to work. An introduction to industrial and organizational psychology* (2. ed.). Chicago: The Dorsey Press.

Mumford, M., Harding, F., Weeks, J. & Fleishman, E. (1988). Relations Between Student Characteristics, Course Content, and Training Outcomes: An Integrative Modeling Effort. *Journal of Applied Psychology*, *73*, 443–456.

Nagano-Saito, A., Leyton, M., Monchi, O., Goldberg, Y. & He, Y. (2008). Dopamine depletion impairs frontostriatal functional connectivity during a set-shifting task. *Journal of Neuroscience*, *28*, 3697–3706.

Nelson, H. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.

Nerdinger, F., Blickle, G. & Schaper, N. (1999). *Arbeits- und Organisationspsychologie.* Berlin: Springer.

Ormerod, T., Richardson, J. & Shepherd, A. (1998). Enhancing the usability of a task analysis method: a notation and environment for requirements specification. *Ergonomics*, *41*, 1642–1663.

O'Connor, P. & Flin, R. (2003). Crew resource management training for offshore oil production teams. *Safety Science*, *41*, 591–609.

Patrick, J. (1992). *Training: Research and Practice.* Amsterdam: Academic Press.

Pintrich, P., Cross, D., Kozma, R. & McKeachie, W. (1986). Instructional psychology. *Annual Review of Psychology*, *37*, 611–651.

Quiñones, M. & Ehrenstein, A. (1997). Preface: Toward a synthesis of psychological research on training. In M. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace. Applications of psychological research* (pp. xi–xii). Washington, D.D.: American Psychological Association.

Rasmussen, J. (1990). Mental models and the control of action in complex environments. In J. Rasmussen (Ed.), *Mental models and human-computer interaction 1* (p. 41–69). North-Holland: Elsevier.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*, 183–213.

Reason, J. & Hobbs, A. (2003). *Managing maintenance error. A practical guide.* Hampshire: Ashgate Publishing.

Rowe, A., Cooke, N., Hall, E. & Halgren, T. (1996). Toward an on-line knowledge assessment methodology: Building on the relationship between knowing and doing. *Journal of Experimental Psychology*, *2*, 31–47.

Salas, E., Wilson, K., Burke, C. & Wightman, D. (2006a). Does crew resource management work? An update, an extension, and some critical needs. *Human Factors*, *48*, 392–412.

Salas, E., Wilson, K., Priest, H. & Guthrie, J. (2006b). Design, delivery, and evaluation of training systems. In E. Salas, K. Wilson, H. Priest

& J. Guthrie (Eds.), *Handbook of Human Factors and ergonomics* (p. 472–512). Hoboken: Wiley.

Salvendy, G. (2006). *Handbook of human factors and ergonomics* (3. ed.). New York: John Wiley & Sons.

Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar Big-Five Markers. *Journal of Personality Assessment*, *63*, 506–516.

Sauer, J., Burkolter, D., Kluge, A., Ritzmann, S. & Schüler, K. (2008). The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics*, *51*, 953–967.

Sauer, J., Hockey, G. & Wastell, D. (2000b). Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics*, *43*, 2043–2064.

Sauer, J., Wastell, D. & Hockey, G. (2000a). A conceptual framework for designing microworlds for complex work domains: a case study of the Cabin Air Management System. *Computers in Human Behavior*, *16*, 45–58.

Saus, E.-R., Johnsen, B. H., Eid, J., Riisem, P. K., Andersen, R. & Thayer, J. (2006). The effect of brief situational awareness training in a police shooting simulator: An experimental study. *Military Psychology*, *18*, 3–21.

Schmidt, F. & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schoofs, D., Preuß, D. & Wolf, O. (2008). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology*, *33*, 643–653.

Schyns, B. & Collani, G. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology*, *11*, 219–241.

Sheridan, T. (2002). *Humans and automation: System design and research issues.* New York: John Wiley & Sons.

Sheridan, T. (2006). Supervisory control. In T. Sheridan (Ed.), *Handbook of human factors and ergonomics* (p. 1025–1052). New York: John Wiley & Sons.

Spiro, R., Feltovich, P. & Coulson, R. (1996). Two Epistemic world-Views: Prefigurative Schemas and Learning in Complex Domains. *Applied Cognitive Psychology*, *10*, 51–61.

Strohschneider, S. (2008). Human-Factors-training. In P. Badke-Schaub, G. Hofinger & K. Lauche (Eds.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen* (pp. 289–306). Heidelberg: Springer.

Szalma, J. (2009). Individual differences in human–technology interaction: incorporating variation in human characteristics into human factors and ergonomics research and design. *Theoretical Issues in Ergonomics Science*, *10*, 381–397.

Thayer, P. (1997). A rapidly changing world: Some implications for training systems in the year 2001 and beyond. In M. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace. Applications of psychological research* (pp. 15–30). Washington, D.D.: American Psychological Association.

Vicente, K. (2006). *The human factor. Revolutionizing the way people live with technology.* New York: Routledge.

Vicente, K., Mumaw, R. & Roth, E. (2004). Operator monitoring in a complex dynamic work environment: a qualitative cognitive model based on field observations. *Theoretical Issues in Ergonomics*, *5*, 359–384.

Weinert, A. (2004). *Organisations- und Personalpsychologie.* Weinheim: Beltz.

Wickens, C., Gordon, S. & Liu, Y. (1998). *An Introduction to Human Factors Engineering.* New York: Addison Wesley.

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

Wickens, C. & McCarley, J. (2008). *Applied attention theory.* Boca Raton: CRC Press.

Wittmann, W. & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science*, *21*, 393–409.

Wonderlic. (2002). *Wonderlic personnel test.* Libertyville: Wonderlic Inc.

Scientific Resume

1984 – 1995	Elementary school in Kyoto, Japan and Mönchaltorf, Switzerland
1995 – 2000	High school in Harstad, Norway and Wetzikon, Switzerland
2000 – 2005	University studies in Psychology, M.Sc. from the University of Zurich in 12/2005
2005 – 2008	Research assistant at the Chair of Organizational Psychology, University of St. Gallen, Switzerland within the research project (SNF) "Cognitive requirement analysis to derive training models for controlling technical systems"
2008 – present	Research assistant at the Institute of Business and Organizational Psychology, Department of Computer Science and Applied Cognitive Science, Faculty of Engineering, University of Duisburg-Essen, Germany

Publications:

- **Burkolter, D.**, Kluge, A., Sauer, J. & Ritzmann, S. (2009). Predictive qualities of operator characteristics for process control performance: The influence of personality and cognitive variables. *Ergonomics, 52*, 302–311. [Included in this doctoral thesis].

- **Burkolter, D.**, Kluge, A. & Brand, M. (2009). Cognitive and personality variables of operators: individual characteristics and process control performance. In D. de Waard, J. Godthelp, F.L. & K.A. Brookhuis (Eds.). *Human factors, security and safety* (pp. 77–88). Maastricht: Shaker Publishing. [Included in this doctoral thesis].

- **Burkolter, D.**, Kluge, A., Sauer, J. & Ritzmann, S. (2010). Comparative study of three training methods for enhancing process control performance: emphasis shift training, situation awareness training, and drill and practice. Manuscript resubmitted for publication in *Computers in Human Behavior, 26*, 976–986. [Included in this doctoral thesis].

- **Burkolter, D.**, Meyer, B., Kluge, A. & Sauer, J. (in press). Assessment of structural knowledge as a training outcome in process control environments. *Human Factors.* [Included in this doctoral thesis].

- Kluge, A., Sauer, J., **Burkolter, D.** & Ritzmann, S. (in press). Designing training for temporal and adaptive transfer. *Journal of Educational Computing Research, 43*(3).

- Kluge, A. & **Burkolter, D.** (in press). Simulatoren und Simulationen als Bestandteil der Fernausbildung. Die St. Galler Simulatorphilosophie und ihre Heuristiken [Simulators and simulations as a constituent of training. The simulator philosophy of St. Gallen and its heuristics]. In M. Schulz & A. Neusius (Eds.), *Fernausbildung schärft Perspektiven – Technologiegestützte Bildung als Motor von Innovationsprozessen.* Augsburg: Ziel-Verlag.

- Kluge, A., Sauer, J., Schüler, K. & **Burkolter, D.** (2009). Designing training for process control simulators: a critical analysis of the research literature. *Theoretical Issues in Ergonomics Science, 6*, 489–509.

- **Burkolter, D.**, Kluge, A., German, S. & Grauel, B. (2009). Waste Water Treatment Simulation (WaTr Sim): Validation of a new process control simulation tool for experimental training research. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 1969–1973). Santa Monica: HFES.

- Kluge, A. & **Burkolter, D.** (2009). *Simulators and simulations: A "systematic approach to training"–point of view.* Research report. University of Duisburg-Essen, Faculty of Engineering, Department of Computer Science and Applied Cognitive Science, Business and Organizational Psychology. August 2009.

- Kluge, A. & **Burkolter, D.** (2008). Training principles for process control tasks and innovative approaches to training. In *Joint International Symposium of ISSNP 2008/CSEP 2008/ISOFIC 2008: Toward innovative nuclear safety and simulation technology, Volume 2* (pp. 99–106). Harbin, China: Harbin University Press.

- Kluge, A. & **Burkolter, D.** (2008). Simulator training for truck driving: indications of learning effects. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid & K.A. Brookhuis (Eds.), *Human factors for assisstance and automation* (pp. 217–227). Maastricht: Shaker Publishing.

- Sauer, J., **Burkolter, D.**, Kluge, A., Ritzmann, S. & Schüler, K. (2008). The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics, 51*, 953–967.

- Kluge, A., Schüler, K. & **Burkolter, D.** (2008). Simulatortrainings für Prozesskontrolltätigkeiten am Beispiel von Raffinerien: Psychologische Trainingsprinzipien für die Praxis. [Simulator trainings for process

control tasks using the example of refineries: Psychological training principles for practice]. *Zeitschrift für Arbeitswissenschaft, 62*, 97–109.

- Kluge, A., Sauer, J., **Burkolter, D.** & Schüler, K. (2007). Konzeption einer Trainingstheorie: Trainingsprinzipien für Prozesskontrolltätigkeiten. [Conceptual design of a training theory: Training principles for process control tasks.] In M. Rötting, G. Wozny, A. Klostermann & J. Huss (Eds.) *Prospektive Gestaltung von Mensch-Technik-Interaktion* (pp. 143–148). 7. Berliner Werkstatt Mensch-Maschine-Systeme (ZMMS Spektrum Band 21). B-Düsseldorf: VDI Verlag.

- **Burkolter, D.**, Kluge, A., Schüler, K., Sauer, J. & Ritzmann, S. (2007). Cognitive requirements analysis to derive training models for controlling complex systems. In D. de Waard, G.R.J. Hockey, P. Nickel & K.A. Brookhuis (Eds.), *Human Factors Issues in Complex System Performance* (pp. 475–484). Maastricht: Shaker Publishing.

- **Burkolter, D.** (2005). Tätigkeitsmerkmale der Freiwilligenarbeit und ihr Zusammenhang mit intrinsischer Motivation und Arbeitszufriedenheit. [Job characteristics of volunteering and their relationship with intrinsic motivation and job satisfaction]. In P. v. Farago & H. Ammann (Hrsg.), *Monetarisierung der Freiwilligkeit. Referate und Zusammenfassungen der 5. Tagung der Europäischen Freiwilligenuniversität* (S. 171–172). Zurich: Seismo Verlag.

- **Burkolter, D.** (2005). *Tätigkeitsmerkmale der Freiwilligenarbeit und ihr Zusammenhang mit intrinsischer Motivation sowie Arbeitszufriedenheit. [Job characteristics of volunteering and their relationship with intrinsic motivation and job satisfaction].* Unpublished master's thesis, Philosophical Faculty, University of Zurich, Switzerland. February 2005.

Talks:

- Burkolter, D., Kluge, A., German, S. & Grauel, B. (2009). Waste Water Treatment Simulation (WaTr Sim): Validation of a new process control simulation tool for experimental training research. Paper presented at the Human Factors and Ergonomics Society 53rd Annual Meeting, San Antonio, TX, USA, 10/2009.

- Kluge, A., Burkolter, D. & Grauel, B. (2009). Analysis of individual performance and interaction with a complex, dynamic task by means of a cognitive error analysis. 6th Conference of the Media Psychology Division (German Psychological Society), Duisburg, Germany, 9/2009.

- Burkolter, D., Kluge, A., Brand, M., Ritzmann, S., & Sauer, J. (2008). Cognitive and personality variables of operators: individual characteristics and process control performance. Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Soesterberg, The Netherlands, 10/2008.

- Burkolter, D. & Kluge, A. (2008). Training principles for process control tasks and innovative approaches to training. Joint International Symposium of ISSNP 2008/CSEP 2008/ISOFIC 2008, Harbin, China, 9/2008.

- Burkolter, D. & Kluge, A. (2008). The effectiveness of emphasis shift training, situation awareness training and drill and practice for system control and fault finding performance in process control. XXIX. International Congress of Psychology, Berlin, Germany, 7/2008.

- Burkolter, D. (2008). Simulatortrainings zur zuverlässigen Steuerung komplexer Systeme. Invited talk at the Northwestern University of Applied Sciences in Brugg, Switzerland, 3/2008.

- Burkolter, D. & Kluge A. (2008). Neues aus der Trainingsforschung für Prozesskontrolltätigkeiten. 11. Erfahrungsaustausch der Simulatortrainer, OMV Raffinerie, Wien Schwechat, Austria, 5/2008.

- Burkolter, D. & Kluge, A. (2007). Evaluating the added value of simulator training for truck driving. Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Braunschweig, Germany, 10/2007.

- Burkolter, D., Kluge, A., Sauer, J., & Schüler, K. (2007). Konzeption einer Trainingstheorie: Trainingsprinzipien für Prozesskontrolltätigkeiten. 7. Berliner Werkstatt Mensch-Maschine-Systeme, Berlin, Germany, 10/2007.

- Burkolter, D., Kluge, A., & Schüler, K. (2007). Cognitive requirements of a process control task for training design: application of task analysis and cognitive task analysis methods. XIIIth European Congress of Work and Organizational Psychology, Stockholm, Sweden, 5/2007.

- Burkolter, D., Kluge, A., & Sauer, J. (2007). Simulation-based training at the workplace: Training needs analysis, training methods and transfer. Chair of Symposium at the XIIIth European Congress of Work and Organizational Psychology, Stockholm, Sweden, 5/2007.

- Kluge, A., Burkolter, D., Schüler, K., Sauer, J. & Ritzmann, S. (2007). Zusammenhang von personbezogenen Variablen und Leistung in der Prozesskontrolle. 5. Tagung der Fachgruppe Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie, Trier, Germany, 9/2007.

- Sauer, J., Ritzmann, S., Kluge, A., Burkolter, D. & Schüler, K. (2007). Das Training heuristischer Regeln in der Prozesskontrolle. 5. Tagung der Fachgruppe Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie, Trier, Germany, 9/2007.

- Schüler, K., Kluge, A. & Burkolter, D. (2007). Drill and practice, error training and process control tasks. XIIIth European Congress of Work and Organizational Psychology, Stockholm, Sweden, May 2007.
- Burkolter, D., Kluge, A., Schüler, K., Sauer, J., & Ritzmann, S. (2006). Cognitive requirements analysis to derive training models for controlling complex systems. Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Sheffield, UK, 11/2006.
- Burkolter, D., Kluge, A., & Schüler, K. (2006). Training principles for simulator training. Berlin Conference on Expertise in Context. Research and Networks for the Study of Expertise, Berlin, Germany, 7/2006.
- Burkolter, D., Güntert, S., & Wehner, T. (2006). Tätigkeitsmerkmale der Freiwilligenarbeit und ihr Zusammenhang mit Arbeitszufriedenheit. 45. Kongress der Deutschen Gesellschaft für Psychologie, Nürnberg, Germany, 9/2006.
- Burkolter, D. (2005). Frei-gemeinnützige Tätigkeit – mehr als "Engagement". Tätigkeitsmerkmale der Freiwilligenarbeit und ihr Zusammenhang mit intrinsischer Motivation und Arbeitszufriedenheit. 5. Tagung der Europäischen Freiwilligenuniversität, Luzern, Switzerland, 5/2005.

Teaching:

- Summer 2009 and 2010: Tutorial, Inferential statistics, Applied cognitive and media sciences (Bachelor), University of Duisburg-Essen, Germany.

- Winter 2008/09 and 2009/10: Seminary, Consumer psychology (deepening), Applied cognitive and media sciences (Master), University of Duisburg-Essen, Germany.

- Summer 2008, Winter 2008/09 & 2009/10: Project course (Development of an advertising concept; consumer research), Applied cognitive and media sciences (Bachelor), University of Duisburg-Essen, Germany.

- Summer 2009: Seminary, Consumer and advertising psychology, Psychology (Bachelor/Master), University of Fribourg, Switzerland (with Prof. Dr. A. Kluge).

- Summer 2008, 2009 & 2010: Lecture, Consumer psychology, Applied cognitive and media sciences (Bachelor), University of Duisburg-Essen, Germany.

- Winter 2007/08: Seminary, Organizational development, Economics (Master), University of St. Gallen, Switzerland (with Prof. Dr. T.S. Eberle, Prof. Dr. C. Steyaert, lic.phil. S. Ritzmann & lic.oec. F. Elliker).

- Winter 2006/07 & 2007/08: Tutorial, Methods of scientific working, Economics (Bachelor), University of Zurich, Switzerland.

- Summer 2006: Seminary, Organizational learning and change, Economics (Bachelor), University of St. Gallen, Switzerland (with Prof. Dr. A. Kluge).

- Supervision of bachelor and master theses, University of Duisburg-Essen, Germany.

Memberships:

2008 – Human Factors and Ergonomics Society

2008 – Human Factors and Ergonomics Society Europe Chapter

Languages:

German, English, Norwegian, French, Japanese, Spanish