Dynamic Task Selection in Aviation Training

Ron Salden

Interuniversitair Centrum voor Onderwijsonderzoek

ISBN: 90-9019113-5
Publisher: Datawyse Maastricht
Cover artwork: Mattias Norén
This dissertation is also published at the National Aerospace Laboratory NLR as publication TP-2004-465.

Dynamic Task Selection in Aviation Training

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit Nederland
op gezag van de rector magnificus
prof. dr. ir. F. Mulder
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 22 april 2005 te Heerlen
om 16:00 uur precies

door
Ron Johannes Carolus Maria Salden
geboren op 22 december 1976 te Geleen

Datawyse

Promotor:
Prof. dr. J.J.G. van Merriënboer, Open Universiteit Nederland

Toegevoegd promotor:
Dr. G.W.C. Paas, Open Universiteit Nederland

Overige leden beoordelingscommissie:
Prof. dr. D. Leutner, Duisburg-Essen University
Prof. dr. J.J.M. van der Sanden, Technische Universiteit Eindhoven
Dr. J. van der Pal, Nationaal Lucht- en Ruimtevaart Laboratorium
Prof. dr. E.J.R. Koper, Open Universiteit Nederland

Datawyse

Acknowledgements

Life can be quite ironic. When fellow students asked me in 1998 if I would see myself as a PhD student after graduation, my answer was a simple yet clear "no way!" Fast forward to October 1st 2000, the date on which I started my very own PhD project. The apparent change of mind came about during my internship in my final college year. The freedom to come up with my own research topic and experimental design was exhilarating. I still hold warm memories of driving around the entire province of Limburg to visit the participants of my experiment. This truly gave new meaning to the phrase 'collecting data'. I am grateful to my internship supervisor Remy Rikers for enabling me to see the wonders of the world of research.

When I found out about a vacant PhD position at the Open University, which was to examine the role of learner characteristics in relation to a rather new concept called 'dynamic task selection', it did not take long for me to try out for this new challenge. Though I must have given an odd impression during the job interview, I was able to convey my enthusiasm about doing research.

My supervisors Jeroen van Merriënboer and Fred Paas had no idea what they had gotten themselves into but were soon to find out. I am much indebted to them for their patience, their unceasing scientific guidance and advising. Though our discussions were relentless at times, it always served the purpose of improving the research in general and more specifically, the acquisition of numerous research skills. Under their guidance I have learned a great deal and have grown into a professional researcher during the course of these 4 years.

In order to be able to conduct experiments one requires participants and I would like to express my gratitude to a few people who have helped me a great deal in this process. For their efforts at the Hogeschool Zuyd many thanks go out to Mr. H.J. Daniels, former head of the technical department (HTS), and to Mr. Léon Sluijsmans, former head of teacher education (PABO). Furthermore, many thanks to Mr. Simon IJspeert from the Aviation department of the Hogeschool Amsterdam for his countless efforts to supply me with participants. Of course I am also much indebted to all the participants who have taken the time to supply me with critical feedback and the desired data sets.

When I arrived at the Open University, the original 'PhD clan' was a rather small but very diverse and comfortable team to join. I am grateful to Liesbeth Kester for fruitful discussions, personal interest and good musical taste; Dominique for being an example of determination and professionalism; Huib for being the ever radiant sun that brightened up the workplace; JW for showing how much work a PhD student can handle if he puts his mind to it and for being the best time-manager ever; Dr. Rob for sharing fanatical movie

interest and to Angela for sharing PhD doubts and reminding me of other important things besides work.

Soon a new batch of PhD students arrived and since that day they keep on coming. Thanks to Silvia for accepting and skillfully mastering the challenge of my only speaking Dutch with her; Tamara for personal interest and being a serious competitor to JW for the 'PhD workaholic of the year' prize; Judith for determination and for joining the giggle team; PJ for keeping up the appearance of being a college student while doing truly innovative research; Gemma for continuing my research and preventing me from feeling like Don Quixote. Also thanks to Karen for doing research in her own way; Marieke for taking the challenge to show these youngsters that age does not matter; Liesbeth Baartman for climbing the walls of the VOP committee; Pieter for taking up the challenge of having the same supervisor as I and to Wendy & Fleuri for supplying the OU with a new dynamic duo.

In addition to the PhD clan there was always a nice niche of young educational technologists who ensured diverse conversation during lunch breaks and our nightly adventures. Thanks to Lieke for showing that one can perfectly work while spending more time on holidays; Frans for musical and political frenzy; Linda for sharing movie interest and cheerfulness; Iwan for musical prowess and entertainment; Owen for being a great rowing partner and proofreading this dissertation and Birgitt for the many late night talks and overwhelming enthusiasm. I am also grateful to Wim van de Vegt for the numerous adaptations to the software of the ATC experiments.

During my PhD project I also got the chance to work and live in Amsterdam for half a year. The National Aerospace Laboratory (NLR) provided me with interesting challenges in a new work environment. Many thanks to my supervisor Jelke van der Pal for the training/science discussions and unwittingly complimenting me on 'thinking too much like a researcher'! Harmen for showing me the value of a good roommate and for accepting me, albeit whilst wearing ear plugs. I loved our conversations just to get my mind off stressful work and the dreadful pressure of deadlines. Thanks to Bart for supplying me with the much needed FMS 'ingredients' for my experiments; Tanja for uplifting humor and for always being up for a night on town. Henrik Schlanbusch who has spent so many hours, day and night, working on Taski for the FMS experiments. Though unfortunately Taski could not used, I am confident it will rise to the occasion and turn some heads in due time! Aemonn for being our VIP spotting guide throughout the numerous allies and bars of Amsterdam; Ellen and Matty for giving NLR a much needed youth injection; Rolf, Mariska, Martin, Hans, Martine and others for creating a good work atmosphere and making lunches pleasant and joyful. Furthermore, thanks to Alexander for the free Pathé cinema visits and not walking out during "Dogville"!

It has been a blessing to have been able to ventilate my frustrations over scientific writing into a much less constrained writing environment. We started our webzine Vampire magazine in 1997 and it always provided me with the opportunity to explore writing in a very different direction than that of science. Yet this direction is one which I am also very passionate about and anyone who knows me to some extent will realize I am talking about music. There are many colleagues who contribute to our webzine but I would like to name a few. Thanks to Ricardo for asking me way back in 1997 to write about our beloved music style and for being able to deal with all my strange Limburgian character traits; Danny for being a brutally honest and down-to-earth guy always trying to persuade me into a game of chess, especially when I *really* did not have the time for it; Joost for deviant taste in music and always providing me with a place to crash when at a loose end in Amsterdam! Past and present members Iwan, Loek, Steven, Geert, Vincent, Tim, Filip, John, Serge, Raymond, Boris, Mark, the list goes on and on. Thanks to all for the great fun moments in real life or on the internet.

I would also like to thank some friends who have given me annual relief from work. No matter if we invaded the Dynamo Open Air festival, the Ardennes or a city like Dublin, the fun factor was extremely high and the alcohol consumed always soared to impressive heights. Thanks to Erik, Peter, Bram, Emiel, Marc, Bert, Lars, Iwan, Stef for their strange, humorous and interesting escapades which always provided me with new wonderful insights. Though, often, I was the only one sober I enjoyed all these occasions. And yes, even the freaking mountain biking trip which nearly killed me!

Special thanks to Mattias Norén for dealing with my long, rambling e-mails filled with crazy ideas, for understanding what I was looking for and for creating the amazing cover artwork which will surely attract people's curiosity.

Last but certainly not least, I would like to thank my 'paranimfen' Steven and Bas who have often lent an ear to my scientific frustrations, helped out with practical stuff and have always been willing to take my mind off of these things. They have been a great support in unfortunate personal matters which I truly value. I am proud to call them my friends.

The future in terms of my career is still clouded, yet I am confident it will clear soon. However, one thing is crystal clear and that is my 2 months recreational visit to Australia and New Zealand. This will provide a much needed break after almost 3 years without a decent vacation. Oceania here I come! To all those I leave behind: Stay tuned folks…

Ron Salden
January 2005

Contents

Chapter 1 – Introduction

Since the emergence of the personal computer and the Internet, new technological developments seem to accelerate with an ever increasing speed. Nowadays, youngsters are seen as Captain Caveman if they are not equipped with their own mobile phone, Personal Digital Assistent with wireless internet connection, and flash-memory MP3 player. These gadgets are exemplary for the ongoing technological developments that ask for continuous adaptation and, indeed, lifelong learning from citizens in modern society. This poses new demands to the field of education in at least two ways. First, it is becoming more important than ever to educate students in such a way that they can deal with rapid changes in their professional and daily-life environment. The transfer of learning to ever changing situations is now far more important than the direct learning outcomes. Second, education must be available to students whenever they need it. The educational research field shows two main tendencies to cope with these new demands: (a) an increasing use of real-life tasks as the driving force for learning, and (b) an increasing flexibility of educational programs as indicated by terms such as "just-in-time-learning" and "education-on-demand".

The first tendency shows that meaningful learning tasks, which are based on real-life tasks, are increasingly used as the "backbone" of educational programs. In the design and development of learning tasks several aspects should be taken into account, which ensure that the tasks are at a suitable level of difficulty for the learners and provide an appropriate amount of support and guidance. Furthermore, the tasks should be authentic, engage and motivate the learners, and make meaningful use of technology. This development is evident in educational approaches such as problem-based learning (PBL), task-oriented learning, and competence-based learning. The basic idea is that the use of real-life tasks helps learners to integrate knowledge, skills, and attitudes into rich cognitive structures, which better allow for transfer of learning as well as new learning in future situations.

The second tendency shows the need to make educational programs more flexible. Life-long learning requires flexible curricula and instructional materials that can be adapted to the needs of the individual learners. People need both a content and a form in education that is directly relevant to their current needs ("education-on-demand"). And they need to be given the possibility to have education at the exact time and place they need it ("just-in-time learning"). Instead of providing the same curriculum to a whole group of students, every student might thus receive a uniquely personalized curriculum.

This dissertation brings both tendencies together. On the one hand, it takes a view on educational programs where a sequence of meaningful learning

tasks serves as the backbone of the whole curriculum. On the other hand, this sequence of learning tasks is conceived as a dynamic entity where each next learning task can be selected in such a way that it best suits the needs of an individual learner.

Research Questions

Flexible learning on the basis of meaningful learning tasks requires some form of dynamic task selection. An intelligent agent (e.g., teacher, training program, trainee) makes decisions about the most optimal learning-task sequence during the training or teaching process. In order to make appropriate decisions, information on the student's progress is used such as indications of the level of performance (e.g., speed, accuracy, errors) and the costs related to reaching this performance (e.g., necessary time-on-task, invested mental effort). The main research question of this dissertation is how dynamic task selection can be used to optimize training programs, the learning process, and transfer test performance. More specific research questions focus on the different types of information that are required to effectively use dynamic task selection and on the role of the trainees themselves in this task selection process. For example, do performance measures contain sufficient information for dynamic task selection or are other measures, such as invested mental effort, also important to take into account? And to what extent are trainees able to fulfill an active role in the process of task selection?

Overview of the Dissertation

In order to answer the research questions, the theoretical framework of the dissertation is given in Chapter 2. Chapters 3 through 5 present four empirical studies on the use and effects of personalized methods in training programs for complex cognitive skills in the aviation domain (i.e., controlling air traffic and programming flight management systems). A closer look is taken at the transfer effects of the mental efficiency method (Paas & van Merriënboer, 1993, 1994a, 1994b), which bases task selection on mental efficiency as a combination of learner's performance and invested mental effort. High mental efficiency is associated with high performance combined with low mental effort, and low efficiency is associated with low performance combined with high mental effort. The mental efficiency method is compared with other personalized training methods (e.g., based only on performance or invested mental effort) and with a non-personalized training method.

Chapter 2 presents a theoretical comparison of approaches to learning task selection that have been used throughout the last three decades in training programs for complex cognitive skills. In general, a development from static part-task selection to dynamic whole-task selection of learning tasks can be

noticed. More specifically, four approaches are identified: (a) a static part-task approach, (b) a static whole-task approach, (c) a dynamic part-task approach, and (d) a dynamic whole-task approach. These four approaches are compared in terms of their flexibility and adaptability to the needs of the individual trainee during training. Furthermore, they are compared with regard to the nature of the complex cognitive skills for which they may or may not be useful training methods.

In Chapter 3, the differential effects of four task selection methods on training efficiency (e.g., training time and number of tasks needed to reach the exit performance level) and transfer test performance are investigated in a computer-based training program for Air Traffic Control (ATC). A non-dynamic condition, in which the learning tasks are presented to the participants in a fixed, predetermined sequence, is compared to three dynamic conditions. The dynamic conditions select learning tasks on the basis of performance, mental effort, or mental efficiency (i.e., a combination of performance and mental effort). The participants are first given an introduction to the ATC field and have to complete a practice task before they start with the actual training program. All participants start with a task of the lowest complexity level and then continue with learning tasks that are selected according to the condition they work in. After the training is completed, they are presented with ten transfer tasks.

Chapter 4 investigates the effects of two personalized training methods on training efficiency and transfer test performance in a computer-based ATC training program. In one personalized condition, task selection is based on a combination of performance and invested mental effort (i.e., mental efficiency); in the other personalized condition, the learner is free to select the complexity level of the next learning task (i.e., learner control). Furthermore, participants in both personalized conditions are matched to "yoked" participants in two control conditions. That is, each individualized training sequence of a participant in the mental efficiency condition or the learner control condition is also presented to a participant in the corresponding yoked control condition. Note that the yoked participant is presented with the training sequence of someone else; hence no personalization occurs in the yoked conditions. After an introduction to the ATC field, all participants are given a pre-training before they start with the actual training program. After completion of the training all participants are presented with a two-fold transfer test consisting of a reaction time test and ten transfer test tasks.

Chapter 5 presents two closely related empirical studies. The first study examines the effects of three task selection methods on training efficiency and test performance in a computer-based training program for programming a Flight Management System (FMS). A non-dynamic condition, in which the learning tasks are presented to the participants in a fixed, predetermined

sequence, is compared to two dynamic conditions. In the dynamic conditions, the learning tasks are either selected by the participants themselves (i.e., learner control) or by a task selection algorithm in the computer-based training program that uses the participant's self-ratings for performance and mental effort. The participants in the learner control condition have total freedom in selecting the learning task they want to practice next. All participants are presented with five test tasks after completion of the training. Since the data from this study suggest that some participants systematically overrate their performance, the role of self-ratings is further investigated in a second study. The non-dynamic fixed condition is again compared to a mental efficiency condition in which students assess their own performance and mental effort. As in the first study, five test tasks are given after the participants have completed the training.

Chapter 6, the final chapter of the dissertation, presents a general discussion of the theoretical framework and the empirical studies. A review of the main results is given, followed by a discussion of the limitations of the conducted experiments. Furthermore, the theoretical and practical implications of the studies are discussed and suggestions for future research are given. The dissertation concludes with some final remarks on the value of dynamic task selection in education.

References

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & van Merriënboer, J. J. G. (1994a). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology, 86*, 122-133.

Paas, F., & van Merriënboer, J. J. G. (1994b). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371.

Chapter 2 – A Comparison of Approaches to Learning Task Selection in the Training of Complex Cognitive Skills [1]

Abstract

This paper presents a comparison of learning task selection approaches that have been used throughout the last three decades in the training of complex cognitive skills. In general, a development from static part-task selection to dynamic whole-task selection can be noticed. The four approaches of static part-task approaches, static whole-task approaches, dynamic part-task approaches, and dynamic whole-task approaches are identified and compared in terms of their flexibility and adaptability to the needs of the individual trainee during training. The comparison shows that dynamic whole-task approaches are the most flexible and adaptive. For each approach it is discussed to what complex cognitive skills they might be useful training methods.

Introduction

Employees are faced with increasingly demanding working environments in modern society. Especially in technical domains such as aircraft control and chemical industry, in which mistakes can lead to dangerous situations and high costs. However, the available training time in which the complex job skills have to be mastered, is limited. The question of how employees can be efficiently trained is considered important. This paper presents a comparison of learning task selection approaches in the training of complex cognitive skills that have been used throughout the last three decades. An important focus of this article is to determine how these approaches can be used to personalize training in order to achieve transferable skills.

A first distinction is made between static and dynamic approaches in the selection of learning tasks. Although both approaches take prior knowledge into account in the development of the training program, the sequence of learning tasks can be determined by the training program or by the trainer either prior to the start of the training, i.e., static approaches, or can be adjusted during the training, i.e., dynamic approaches. Furthermore, both approaches are subdivided into part- and whole-task approaches. This second distinction reflects the development of training programs from static part-task based to dynamic whole-task based. First, the static part-task and whole-task approaches are discussed that are characterized by a preset order and complexity of learning tasks prior to the training. Then, dynamic part-task and whole-task

[1] *This chapter is currently in press as: Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (in press). A comparison of approaches to learning task selection in the training of complex cognitive skills. Computers in Human Behavior.*

approaches are discussed that are characterized by the possibility to adjust the order and complexity of learning tasks during training.

The four approaches are compared in terms of their flexibility and adaptability to the needs of the individual trainee during training, using the following factors: clear determination of learning tasks, no integrative constraints, ability to adjust during training, personalized instruction, possibility to use cognitive load for determining task selection and coping with high task organization.

For a task selection approach to be useful it should be able to determine a set of clear-cut learning tasks. Furthermore, it should be able to integrate parts of a whole-task with ease and deal with parts that are related to each other (high task organization). In order to achieve personalized training, the focus should be on the individual student instead of on a group of students. The ability to adjust during training enables one to alter the complexity and order of learning tasks when a student encounters a problem. While usually performance measures are used to determine task selection, the concept of cognitive load is getting more and more acknowledgement as an important factor to take into account as well (e.g., Brusilovsky, 1992; Kashihara, Hirashima, & Toyoda, 1995). Finally, the results from the comparison of the four learning task selection approaches will be used to discuss for what complex cognitive skills they might be useful training methods.

Static Part-Task Selection Approaches

Part-task approaches were originally proposed because it was considered impossible to start training with learning tasks that represent the full complexity of the authentic task. A learner's cognitive system might be overloaded, which can negatively affect learning, performance and motivation (Sweller, van Merriënboer, & Paas, 1998). Furthermore, Wightman and Lintern (1985) have proposed that part-task training can have higher learning efficiency and lower training costs than whole-task training, especially, when task complexity is high and task organization is low. Task complexity refers to the load that each separate component of the task imposes on the learner's cognitive system, while task organization refers to the processing demands of the interacting components of the task (Fabiani, Buckley, Gratton, Coles, Donchin, & Logie, 1989).

Besides a preset order and complexity of learning tasks, which is determined prior to the training, static part-task approaches are characterized by the fact that all learning tasks include some of the skills that a learner should acquire and be able to apply after the completion of the training. Learners start with part-tasks and practice increasingly larger parts until they have mastered the whole-task.

Part-tasks can consist of segments, fractions, simplifications or prerequisites of the whole-task. If a task contains subtasks that can be clearly separated, then the whole-task can be segmented into series of these subtasks (Wightman & Lintern, 1985). For example, when learning to drive a car, skills such as steering, accelerating, and shifting gear are three possible subtasks. Fractionation breaks elements of the whole-task that are normally performed concurrently into components. For example, trainees in driving are typically required to learn how to steer a car before they can continue to use the pedals and the shifting gear. Simplification implies the reduction of difficulty of one or more elements of the task. For example, one will first practice steering a car on a spacious parking lot, before advancing to a more complex, less spacious environment. Prerequisites are parts that need to be acquired before other parts of the task can be mastered. Examples are steering a car, using the pedals, or using the shifting gear. In partitioning a whole-task, the focus is on defining a large number of small tasks in order to yield a so-called fine-grained decomposition of the task. For example, car engineers divide the whole-task of checking an engine into specific parts of the engine. By systematically checking each specific part of the engine they can exclude possible causes of the failing of the engine.

Several approaches may be used to determine the order of the different parts in part-task training. In backward chaining the last component of a task is practiced first and earlier components are introduced later in the training. In forward chaining the order for adding task components is from first to last (Proctor & Dutta, 1995). When considering driving a car, one could practice the shifting gear as part of a backward chaining part-task, and end with practicing how to start the car. In forward chaining one would practice part-tasks in the exact opposite order. Several approaches determine the reintegration of the parts into the whole-task. In repetitive part training each component is practiced separately and then additional parts are added sequentially. In pure part training each component is practiced in isolation before the parts are combined. Progressive part training is a combination of repetitive part and pure part training as each part is practiced in isolation before being added one at a time to the task (Proctor & Dutta, 1995). Finally, snowballing resembles forward chaining in that it starts with offering parts and more parts are added with each next step (Landa, 1983). Figure 1 depicts these sequencing approaches.

Another part-task training approach that determines the order of the parts is the hierarchical approach, which was developed by Gagné (1968). It is based on the observation that a skill is made up of prerequisites or enabling skills that must be learned before the larger, more complex skills, of which they are a part, can be learned. Gagné distinguishes several increasingly detailed and difficult skills whose hierarchical arrangement helps to figure out what prerequisites a given skill might have. To make sure the learner is not

confronted with learning tasks of skills that are already mastered; the training needs to be started at the level of "entering knowledge" of the learner. A hierarchical sequence is one that never teaches a skill before its prerequisites (Reigeluth, 1999).

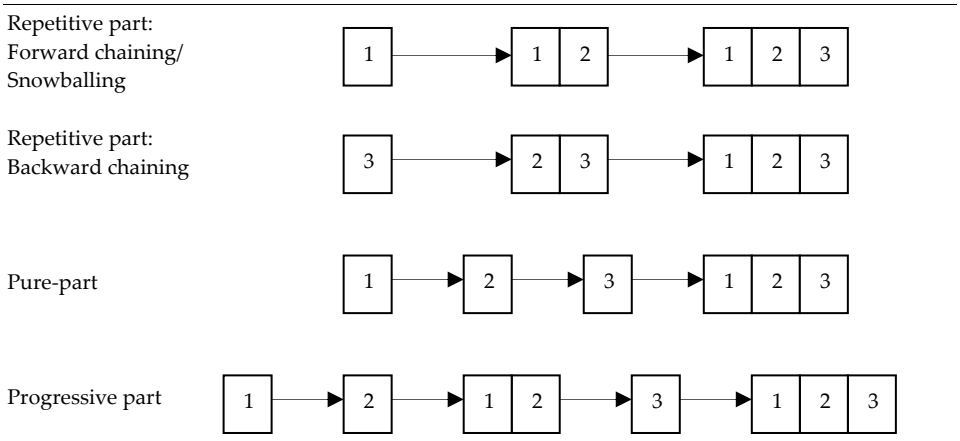

Figure 1. Adapted from Proctor and Dutta (1995): sequencing of a task divided in three serial part-tasks.

Although the discussed approaches have several differences with regard to the preferred application in specific training contexts, they all claim that one should adapt the training to the trainee's prior knowledge and take the growing amount of acquired knowledge of the trainees into account. And like the hierarchical approach states, some skills should be learned before a trainee can start to learn a more complex skill.

However, for complex cognitive tasks, determining the part-tasks is not easily done as many parts are related to each other. Training on parts of a whole-task does impose integrative constraints since a trainee is eventually presented with a whole-task version of all the previous part-tasks. Since the task order and complexity of the part-tasks are preset prior to the training, there is no possibility to make adjustments during training. Furthermore, the training methods focus on a group of students instead of presenting personalized instruction. Though cognitive load (e.g., Sweller, 1989) could be used to determine the order of learning tasks, this could only be done prior to training. Lastly, part-task methods are unable to cope with task organization when parts interact highly with each other (see Table 1).

Static Whole-Task Selection Approaches

Static whole-task selection approaches started to develop when the limitations of the part-task approaches became apparent. Besides a preset order and complexity of learning tasks prior to the training, static whole-task approaches are characterized by the fact that every learning task includes all the skills that a learner should have acquired and be able to apply after the training.

Table 1 *Overview of strengths and limitations of task selection approaches.*

	Static		Dynamic	
	part-task	whole-task	part-task	whole-task
Clear determination of learning tasks	-	+	-	+
No integrative constraints	-	+	-	+
Ability to adjust during training	-	-	+	+
Personalized instruction	-	-	+	+
Possibility to use cognitive load for determining task selection	+	+	+	+
Coping high task organization	-	+	-	+

Reigeluth and Stein's elaboration theory (1983) determines the order and complexity of learning tasks prior to training by means of identifying task expertise. An approach for building task expertise that offers guidance for analyzing, selecting, and sequencing the learning tasks, is the Simplifying Conditions Method (SCM). It claims that, given that any complex task has some conditions under which it is easier to perform than under others, one should start with the simplest version of the task that is still fairly representative of the task as a whole. For example, when learning to drive a car, one will first practice driving on roads not crowded with traffic, before advancing to a more complex, more crowded environment. The SCM gradually progresses to more complex versions of the task until the desired complexity level is reached, making sure that the learner is aware of the relationships between the different task versions. These different versions of learning tasks constitute a task class when they have an equal complexity. Differences in complexity only exist between task classes (van Merriënboer, 1997) and each task class contains learning tasks that are complete, real-world performances of a whole-task (Reigeluth, 1999).

A relatively new approach, which resembles some aspects of the elaboration theory to a high extent, is the familiarity approach (Scheiter, Gerjets,

& Tack, 2001). This approach uses the prior knowledge (familiarity) of the trainees and the difficulty of the tasks to base the training sequence on. The first lessons or parts of a training contain high familiarity aspects and are of low difficulty. As a learner progresses through the lessons or training, familiarity decreases and difficulty increases. For example, when novice learners start learning physics, they are first presented with the basic elements and rules. Only after having acquired these, they will proceed to more complex relationships of the properties of physics. In short, the familiarity approach states that the amount of presented unfamiliar information and complexity of the learning tasks is adjusted to the learner's current knowledge state prior to the training.

A mental model progressions approach states that one should start with learning tasks that require a mental model that contains the ideas that are most simple, representative, fundamental, and concrete (van Merriënboer, 1997). For example, when learning to drive a car one should first learn to perform the basic actions (e.g., steering, shifting gear, and using the pedals) in a spacious environment. The model has to generate tasks that learners can work on by means of taking the prior knowledge of the learner during the first learning tasks into consideration. Progressions can occur in the development of a particular model as well as by changes in model order and degree of elaboration. While a subsequent model adds complexity or detail to an aspect of the former models and becomes an elaboration of them, learners proceed to more complex tasks accordingly to model development. Furthermore, a subsequent model can provide alternative strategies on solving problems in the domain. This process continues until the desired exit behavior is attained, which is specified in a certain level of elaboration and a set of mental models that offer different strategies and perspectives. The general idea is that each subsequent model should allow the learner to solve a new task class (van Merriënboer, 1997). For example, this approach has been used to explain the design and troubleshooting of electrical circuits where learners progress from learning tasks that present a basic idea of a circuit to tasks that present the laws of electricity (White & Frederiksen, 1989).

The emphasis manipulation approach (Gopher, Weil, & Siegel, 1989) was invented to avoid the difficulty of dividing a task into parts and states that learners should be exposed to the whole-task in its full complexity throughout the training period. However, different sets of skills are emphasized during different training phases. The learners are enabled to focus on specific aspects without loosing sight of the whole-task. By emphasizing and de-emphasizing aspects of the whole-task, learners learn to monitor priorities and to direct attention to the changes of emphasis. It is proposed to emphasize skills that are difficult and demanding for the learners and that are sufficiently different from each other. This is expected to lead to changes in performance for the whole-

task when being applied. For example, novice tennis players might focus on their backhand during a training game. When the backhand is mastered, it can be de-emphasized and the service could be emphasized.

All approaches rightfully claim that the training should be adjusted to the prior knowledge of the trainee first. The complexity of the tasks is gradually increased in relation to the growing amount of acquired knowledge of the trainees. However, the emphasis manipulation approach exposes learners to tasks in its full complexity throughout training. This approach does not start the training with the simplest whole-task but rather with a whole-task in which particular aspects of the task are emphasized.

Since the approaches focus on whole-tasks, they can determine the learning tasks quite well and they also experience no integrative constraints. The static nature of the whole-task approaches shows similar limitations as the static part-task approaches. While the order and complexity of learning tasks are set prior to the training, the whole-task approaches lack the ability to make adjustments during training. Furthermore, due to their focus on a group of students, differences between trainees are not taken into account. Also, cognitive load could be used to determine the order and complexity, though only prior to training. Finally, these approaches can cope with high task organization since they present tasks in its entirety during training (see Table 1).

Dynamic Part-Task Selection Approaches

As the role of the computer increased significantly, training programs started to become computer-based. The training programs still used a part-task approach but instead of being static, they became dynamic. This means that it became possible to make adjustments in the order and complexity of the learning tasks during the training phase. The parts and sequencing strategies that these dynamic approaches use are similar to those used by the static part-task approaches.

The first dynamic task selection approach, in a very raw version, was branching (e.g., Coulson, Estavan, Melaragno, & Silberman, 1961; Gilman, 1969). This part-task approach attempts to diagnose to what extent the learner has acquired the skills presented in the training. Snowballing (see static part-task approaches section) is used to reintegrate parts to the whole-task. These parts can be any of the different formats of part-tasks that were discussed in the section on static part-task selection approaches.

Occasionally, when a certain amount of skills has been trained, additional part-tasks are presented to test the learners' performance. If they perform correctly, they branch to the next skill; if they perform incorrectly, they are branched to additional skills training, depending on the mistake they made (Clark, 1997). The amount of branching may vary considerably from occasional

branch points to branching after every learner's response. The direction of branching can either be forward, sideways, or backward. Forward means that the learner skips skills, backward means that the learner repeats skills, and sideways means that the learner is exposed to extra skills training (Allesi & Trollip, 1991).

A part-task program that takes the dynamic aspect one step further than the typical branching is Basic Instructional Program (BIP). BIP is a program for teaching introductory programming (Bar, Beard, & Atkinson, 1976) which attempts to individualize the sequence of instruction through the appropriate selection of part-tasks from a database of learning tasks. These part-tasks involve a varying set of skills that the learners have to acquire. The task selection is based on information contained in a network that relates learning tasks in the training program to issues in the knowledge domain. The training program distinguishes three conceptual layers: techniques, skills, and tasks. These three layers are used to offer learners learning tasks that can include varying skills and techniques (Bar et al., 1976).

BIP uses a student model, which is being updated during training after the completion of each learning task. Selection of the learning tasks is determined on the basis of the student model. New learning tasks are selected if the current task is mastered and does not contain skills that lie beyond the learner's reach (Bar et al., 1976). New skills are added to each new learning task if a learner has mastered a certain set of skills sufficiently. In other words, snowballing is used to reintegrate parts into the whole-task.

In line with the static part-task selection approaches, both branching and BIP rightfully claim that training should be adapted to the trainee's prior knowledge and the growing amount of acquired knowledge of the trainees. Furthermore, some skills should be acquired before a trainee can start learning a task with more complex skills.

Since these approaches focus on part-task training, they have several similar disadvantages as the static part-task approaches. Determining the learning tasks is not easy, especially when task organization is high. However, the advantage of their dynamic nature not only allows them to make adjustments in task order and task complexity during training, it also enables them to actually use cognitive load for task selection during training. Furthermore, following this dynamic nature, a shift in focus occurred from group-based training to personalized instruction (see Table 1).

Dynamic Whole-Task Selection Approaches

The introduction of computer-based training also enabled the use of dynamic whole-task approaches in training complex cognitive skills. While presenting learners with whole-tasks during the training it is also possible to adapt more

efficiently to the needs of the individual trainee. Adjustments can be made in the order and complexity of the learning tasks. The program can respond to the learner's problems during the training, with decisions being made that are typically based on the performance of the trainee.

Like the part-task program BIP (Bar et al., 1976), many Intelligent Tutoring Systems (ITS) use a student model in order to keep track of the trainee's history of the tasks and the corresponding performance. A student model builds a knowledge base of the trainee, and updates that knowledge base as the trainee progresses through the learning tasks. The progress of the trainee is checked on the basis of comparing the trainee's performance to the learning objectives that were specified prior to training. After this comparison, the selection rules indicate the next learning task to present to the learner.

However, many ITS focus on elaborating the operationalization of student modeling while not being clear on the selection rules that are used. These approaches include psychometric approaches (for a discussion, see Everson, 1995), agents (e.g., Capuano, Mersella, & Salerno, 2000; Giroux, Leman, & Marcenac, 1995), and fuzzy logic (e.g., Virvou, Maras, & Tsiriga, 2000). Though an important student models' function is to give specific feedback to the learners about their performance, only a few training approaches exist that explicitly describe the task selection rules being used.

One of these training approaches is the Completion Assignment Constructor (CASCO). CASCO is an ITS that dynamically selects learning tasks in a training of introductory programming (van Merriënboer, Luursema, Kingma, Houweling, & De Vries, 1996). The task selection rules that CASCO uses are straightforward. The most important rule states that a good learning task is suitable to present new learning elements and to practice known learning elements. CASCO can therefore be classified as a Progressive Mental Model (PMM). The other rules state that a good task is not too difficult, has not been presented to the learner before, and is suitable to remediate learning elements the learner makes mistakes with. While learners are working on the learning tasks, learner diagnosis takes place in order to update the student model. Fuzzy logic is being used to operationalize the student modeling. Fuzzy sets are used to keep track of the learner's progress and expertise in order to optimize the selection of the next learning task (van Merriënboer, Krammer, & Maaswinkel, 1994).

Another approach that explicitly describes the task selection rules that are being used is the mental efficiency approach (Paas & van Merriënboer, 1993). While most ITS-programs only use prior knowledge and performance data to determine task selection, the mental efficiency also uses the associated cognitive load (e.g., Kalyuga, Chandler, & Sweller, 1998; Paas & van Merriënboer, 1993). The concept of cognitive load has been acknowledged as an important factor in the training of complex cognitive skills (e.g., Sweller, 1989;

Sweller, et al., 1998). Using the associated cognitive load on a learning task makes sense as learners can achieve a certain performance with varying amounts of cognitive load. It enables one to differentiate between a learner who needed to work laboriously to attain a certain performance level and a learner who attained the same performance level with low mental effort. Whereas the first should certainly not yet be confronted with more complex learning tasks, the latter should be ready to deal with more complex learning tasks.

The task selection rules used in the mental efficiency approach are fairly straightforward. While both performance and cognitive load are measured on a 5-point scoring scale, it is the difference between these two variables that determines the increase or decrease in complexity of the next learning task. One should have a database of learning tasks that are divided over a number of complexity levels. A student who attains a performance score of 4 while his cognitive load is 3, will be presented with a learning task that is one complexity level higher than the previous task. A first indication of the beneficial effects of the mental efficiency approach was found in a study by Camp, Paas, Rikers and van Merriënboer (2001), who used the efficiency method in an Air Traffic Control training.

The focus on whole-tasks of these dynamic approaches resembles several advantages of the static whole-task approaches. First, the dynamic whole-task approaches also state that training should be adjusted to the prior knowledge of the trainee, and that the complexity of the tasks should be increased in relation to the growing amount of acquired knowledge of the trainees. Furthermore, the learning tasks are easily determined and no integrative constraints occur, hereby coping well with high task organization. The dynamic nature of these whole-task approaches also has the advantages of the dynamic part-task approaches. Besides making adjustments in task order and task complexity, one can base these also on cognitive load measures during the training (see Table 1).

Discussion

The comparison has shown that part-task approaches were proposed because it was considered impossible to start training with highly complex learning tasks. This would overload a learner's cognitive system and lead to negative effects on learning, performance and motivation (Sweller et al., 1989). However, determining useful parts from a whole-task proved to be quite difficult (Gopher et al., 1989). Furthermore, part-task approaches cannot easily account for the integrative aspects of complex tasks, which can be very inefficient when time or integrative constraints are high (van Merriënboer, Kirschner, & Kester, 2003). Finally, static part-task approaches are not able to make adjustments during training. Overall, static and dynamic part-task approaches do not cope well

with high task organization and do not have the same ability to use cognitive load to determine task selection as whole-task approaches.

Whole-task approaches can more easily cope with a high task organization because they take the coordination and integration of the parts of the whole-task into account from the beginning of the training. The static whole-task approaches can be considered too inflexible, as they do not allow intervention during training when a learner encounters a problem. Rather, they focus on adapting training to the prior knowledge and growing amount of acquired knowledge of a target group before the training starts. In order to have an optimal learning process the trainer or the training program should be able to make adjustments on those specific moments during training when a learner is faced with a learning task that is too complex to solve at that moment.

Although the dynamic whole-task approaches have extended their capacity to adjust to the needs of the individual and can deal with a high task organization, it is believed that they still miss an important aspect of the learning process, namely the associated cognitive load. Although cognitive load was sometimes measured in dynamic whole-task approaches (e.g., Brusilovsky, 1992; Kashihara et al., 1995), it was never used as a determinant for task selection. While recent studies (Camp et al., 2001; *Chapter 3:* Salden, Paas, Broers, & van Merriënboer, 2004) have investigated the beneficial effects of cognitive load, more research is needed to fully explore its possibilities.

Though each of the four approaches has its limitations, they can still be useful for the training of certain cognitive skills. The static part-task approaches can be used well for training less complex skills with a low task organization. The execution of routines in aircraft maintenance can be trained using a static part-task approach. For example, the order in which the certain particles of an engine are checked is constantly the same. Furthermore, one can isolate the engine particles rather well for inspection.

Dynamic part-task approaches can be used for a training of complex skills that learners cannot start to practice in a whole-task format. For example, when learning to drive a car, part-tasks like steering or shifting gear are practiced first before a student can continue to a more integrated whole-task practice of driving the car. Dynamic part-task approaches can also be incorporated into dynamic whole-task methods. For example, when learning to drive a car, one might perform the part-task of shifting gear not adequately. The trainer might decide to focus on this part-task before the student can continue with practicing the whole-task of driving the car.

Static whole-task approaches can be used for a training in which the tasks are performed in a specific order. For example, when a physician is diagnosing a patient, he or she will follow a certain standard procedure. After initial interviews with the patient, the physician determines what physical examinations have to be performed. Then, the physician studies the results and

is able to minimize the possible diagnoses that might apply to the patient's case. After further interviewing of the patient a final diagnosis is made and the treatment is determined.

The dynamic whole-task approaches can be used for a wide range of training programs to learn complex skills because of their highly flexible and adaptive nature. For example, the Air Traffic Control (ATC) domain is exemplary for complex cognitive skills in which task organization is high and the cognitive system of the learner is highly imposed. Instructions can be personalized for each individual student, and if a student should encounters a problem, one can respond flexible and adapt the material and instructions to the student's performance and cognitive load measures during the training.

In conclusion, to attain efficient instructional methods, it is important to adapt instruction to the individual learner. The approaches that were developed over the last three decades have gradually increased the personalization of the training material. Despite the fact that the discussed approaches have their limitations, they can still be useful to train cognitive, complex skills. Also, the combination of some approaches can yield larger benefits in developing an efficient training.

References

Allesi, S. M., & Trollip, S. R. (1991). *Computer-based instruction: Methods and development* (2nd ed.). Englewood Cliffs, NY: Prentice-Hall.

Bar, A., Beard, M., & Atkinson, R. C. (1976). The computer as a tutorial laboratory: the Stanford BIP project. *International Journal of Man-Machine Studies, 8*, 567-596.

Brusilovksy, P. L. (1992). A framework for intelligent knowledge sequencing and task sequencing. In C. Frasson, G. Gauthier, & G. I. McCalla (Eds.), *Second International Conference On Intelligent Tutoring, ITS '92*, (pp. 499-506). Berlin, Germany: Springer-Verlag.

Camp, G., Paas, F., Rikers, R. M. J. P., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Capuano, N., Marsella, M., & Salerno, S. (2000). ABITS: An Agent Based Intelligent Tutoring System for distance learning. In C. Peylo (Ed.), *Proceedings of the International Workshop on Adaptive and Intelligent Web-based Educational Systems held in Conjunction with ITS 2000 Montreal, Canada*. Osnabrück, Germany: Technical Report of the Institute for Semantic Information Processing.

Clark, D. (1997). *Instructional system design -Development phase - Chapter IV*. Retrieved January 18, 2000, from http://www.nwlink.com/~donclark/hrd/sat4.html

Coulson, J. E., Estavan, D. P., Melaragno, R. J., & Silberman, H. F. (1961). *Effects of branching in a computer controlled auto-instruction device* (Report No. TM-617). Santa Monica, LA: System Development Corporation. (ERIC Document Reproduction Service No. ED 002 306)

Everson, H. T. (1995). Modeling the learner in intelligent tutoring systems: The promise of a new psychometrics. *Instructional Science, 23*, 433-452.

Fabiani, M., Buckley, J., Gratton, G., Coles, M. G. H., Donchin, E., & Logie, R. (1989). The training of complex task performance. *Acta Psychologica, 71*, 259-299.

Gagné, R. M. (1968). Learning hierarchies. *Educational Psychologist, 6*, 1-9.

Gilman, D. A. (1969). *Remedial and review branching in computer-assisted instruction* (Report No. TM-617). Santa Monica, LA: System Development Corporation. (ERIC Document Reproduction Service No. EJ 006 338)

Giroux, S., Leman, S., & Marcenac, P. (1995). *Representing organisational learner models: A generic concurrent coordination*. Retrieved November 29, 2000, from http://advlearn.lrdc.pitt.edu/its-arch/papers/leman.html

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

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.

Kashihara, A., Hirashima, T., & Toyoda, J. (1995). A cognitive load application in tutoring. *Journal of User Modeling and User-Adapted Instruction, 4*, 279-303.

Landa, N. L. (1983). The algo-heuristic theory of instruction. In C.M. Reigeluth (Ed.), *Instructional design theories and models* (pp. 163-211). Hillsdale, NJ: Lawrence Erlbaum.

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. American Psychology Texts: Sage Publication.

Reigeluth, C. M. (1999). The elaboration theory: audience for scope and sequence decisions. In C. M. Reigeluth (Ed.), *Instructional design theories and models Vol. II: A new paradigm of instruction theory* (pp. 425-453). Mahwah, NJ: Lawrence Erlbaum.

Reigeluth, C. M., & Stein, R. (1983). Elaboration theory. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status*. Hillsdale, NJ: Lawrence Erlbaum.

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science, 32*, 153-172.

Scheiter, K., Gerjets, P., & Tack, W. H. (2001). *The impact of problem order: sequencing problems as a strategy for improving one's performance*. Unpublished manuscript, Saarland University, Saarbruecken, Germany.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 4*, 457-466.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. Educational *Psychology Review, 10*, 251-296.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educational Psychologist, 38*, 5-13.

Van Merriënboer, J. J. G., Krammer, H. P. M., & Maaswinkel, R. M. (1994). Automating the planning and construction of programming assignments for teaching introductory computer programming. In R. D. Tennyson (Ed.), *Automating instructional design, development, and delivery* (NATO ASI Series F, Vol. 119, pp. 61-77). Berlin: Springer Verlag.

Van Merriënboer, J. J. G., Luursema, J. J., Kingma, H., Houweling, F., & De Vries, A. P. (1996). Fuzzy logic instructional models: The dynamic construction of programming assignments in CASCO. In R. D. Tennyson & A. E. B. Barron (Eds.), *Automating instructional design: Computer-based development and delivery tools* (NATO ASI Series F, Vol. 140, pp. 265-302). Berlin: Springer Verlag.

Virvou, M., Maras, D., & Tsiriga, V. (2000). *Learner modelling in an Intelligent Tutoring System for the passive voice of English language*. Retrieved November 23, 2000, from http://ifets.gmd.de/periodical/vol_4_2000/virvou.html

White, B. Y., & Frederiksen, J. R. (1990). Causal models progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 42*, 99-157.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27*, 267-283.

Chapter 3 – Mental Effort and Performance as Determinants for the Dynamic Selection of Learning Tasks in Air Traffic Control Training [1]

Abstract

The differential effects of four task selection methods on training efficiency and transfer in a computer-based training for Air Traffic Control were investigated. A non-dynamic condition, in which the learning tasks were presented to the participants in a fixed predetermined sequence, was compared to three dynamic conditions, in which learning tasks were selected on the basis of performance, mental effort, and a combination of both (i.e., mental efficiency). Using the 3-factor mental efficiency formula of Tuovinen and Paas (2004), the hypothesis that dynamic task selection leads to more efficient training than non-dynamic task selection was confirmed. However, the hypothesis that dynamic task selection based on mental efficiency leads to more efficient training than dynamic task selection based on performance or mental effort alone was not supported. The results are discussed in light of the theoretical framework and suggestions are given for future research.

Introduction

Within the aviation domain there is a serious shortage of well-trained air traffic controllers, mainly due to the yearly increasing crowdedness of the airspace (Galster, Duley, Masalonis, & Parasuraman, 2001). Relieving the workload of air traffic controllers by using Free Flight (FF) and increasing the efficiency of Air Traffic Control (ATC) training are the two main perspectives that have been put forward as possible solutions to this problem. FF aims at minimizing ATC restrictions by allowing user-preferred routing and free maneuvering (RTCA, 1995). With an advanced level of FF, the role of the air traffic controllers would become less demanding as their primary activity would be to monitor the FF actions (Galster et al., 2001). However, human monitoring of automated systems can be poor, especially if the operator has little active control over the automated process and is engaged in other tasks (Parasuraman, Molly, & Singh, 1993). Furthermore, high levels of automation might cause a loss of traffic awareness in the air traffic controller, which leads to an increase of required time for recovery from a failure.

The second perspective of improving the efficiency of the training of air traffic controllers seems more promising and is considered in this study. One of the main characteristics of complex domains such as ATC is that each task often contains new elements compared to the previous tasks. In other words, each new task can be considered as a transfer task in which the previously acquired

---

[1] *This chapter is published as: Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. Instructional Science, 32, 153-172.*

knowledge needs to be applied differently. One should note that besides new elements, each learning task contains the basic skills that have to be acquired (e.g., giving headings and altitude commands). Though the variability and complexity of the learning tasks increase during training, each task builds upon this basis.

An efficient training offers trainees a powerful learning environment in which they can acquire skills quickly and adequately, and learn how to apply these skills flexibly in new situations and tasks. The non-dynamic instructional methods that are currently being used in ATC training programs do not work efficiently as a large amount of students does not complete the training (EATMP Human Resources Team, 2001). The current article compares three dynamic training task selection methods with regard to their effects on training efficiency.

During the last three decades, training methods and programs have evolved in three important ways (for an overview see *Chapter 2:* Salden, Paas, & van Merriënboer, in press): from static to dynamic, from part-task based to whole-task based, and from group-based to personalized. Especially, the use of personalized selection of learning tasks is believed to be strongly related to increased training efficiency (*Chapter 2:* Salden et al., in press). Although many Intelligent Tutoring Systems (ITS) have extended their capacity to adapt the selection of learning tasks to the individual learner's needs by incorporating student models that keep track of a student's performance history, we claim that they are lacking an important aspect of the learning process, namely, cognitive load. Although, the concept of cognitive load is sometimes measured (e.g., Kashihara, Hirashima, & Toyoda, 1995) it has never been used in ITSs as a determinant for task selection. There is no doubt that cognitive load is a crucial factor in the training of complex cognitive skills (e.g., Sweller, 1989; Sweller, 1999; Sweller, van Merriënboer, & Paas, 1998), but usually, only performance measures such as speed and accuracy are used to select learning tasks.

From the viewpoint of cognitive load theory (Paas, Renkl, & Sweller, 2003), dynamic task selection can be superior to fixed task selection as it provides the possibility to adjust the training to the cognitive state of the learner, thereby controlling the load that is imposed on a learner's cognitive system. Although individual measures of performance and mental effort can be used as indicators of the cognitive demands a certain task places on the learner, the combination of both measures is considered a superior estimate of the cognitive demands in the dynamic selection of training tasks. It is quite feasible for two people to attain the same performance levels, while one of them experiences a very high cognitive load and needs to work laboriously through a very effortful process, whereas the other person experiences a low cognitive load and reaches the same performance level with a minimum of effort. However, most people would agree that the next learning task should be less

difficult for the first person than for the second person. Our claim is that task selection, and consequently training efficiency can be improved by taking the combination of performance and cognitive load measures into account. To obtain a good indication of the cognitive load that is imposed on a person's cognitive system, mental effort measurements are used.

A combined measure of performance and mental effort has been proposed as a measure of mental efficiency by Paas and van Merriënboer (1993; see also Paas, Tuovinen, Tabbers, & Van Gerven, 2003). These authors present a calculational approach for combining measures of mental workload and task performance that allows one to obtain information on the relative efficiency of instructional conditions. It is proposed that learners' behavior in a certain learning condition is more efficient if (1) their performance is higher than might be expected on the basis of their invested mental effort, and/or (2) their invested mental effort is lower than might be expected on the basis of their performance. Thus, a high performance combined with a low mental effort is most efficient and a low performance combined with high mental effort is least efficient. Recently, Tuovinen and Paas (2004) have proposed a new version of the efficiency formula, in which training efficiency is calculated on the basis of three dimensions. The current study adopts this 3D efficiency formula and uses training effort and training time to express the costs associated with training, and test performance to express its benefits.

A first confirmation for the claim that the use of mental efficiency makes the individual training more efficient and leads to better transfer results was found in a study conducted by Camp, Paas, Rikers, and van Merriënboer (2001). They compared four methods of task selection in the ATC domain. In the first method, tasks were presented in a fixed, predetermined sequence from simple to complex. In the other three methods, the tasks were presented dynamically, based on either performance, mental effort, or the combination of both, which is mental efficiency. Results showed that dynamic task selection leads to more efficient training than non-dynamic task selection. Furthermore, dynamic task selection based on mental efficiency did not lead to more efficient training and better transfer than dynamic task selection based on performance or mental effort alone.

The current experiment is a partial replication of Camp et al.'s study (2001). The same three learner variables are used for dynamic task selection. These variables, mental effort, performance and mental efficiency, are used to dynamically determine task complexity in training in the ATC domain. As in the study of Camp et al. cognitive load is measured using a five-point subjective rating scale on which the participants have to indicate their invested mental effort for each task. Performance is measured as the accuracy with which participants guide aircraft to a certain goal location in dynamic ATC-situations. The mental effort and performance scores are combined with the mean total

training time of the conditions to determine the efficiency of the instructional conditions.

Two main differences in the procedure of this study and the Camp et al. (2001) study can be distinguished. First, the present study applies slightly different measurement scales and another selection algorithm. While the scales used by Camp et al. were rather rough because they applied exactly the same performance and mental effort scales for all tasks within the same complexity level, the scales in the current study are sensitive to the differences between the separate learning tasks within the same complexity level. With regard to the selection algorithm that is used for selecting a new learning task of a certain complexity level, the maximum jump size between complexity levels was decreased from four in the Camp et al. study to two in the present study, forcing a smoother increase or decrease in task complexity.

Secondly, a different method of determining the efficiency of the training conditions is used. Three methods can be identified for this goal. The first method investigates which instruction leads to the highest training performance combined with the lowest mental effort during training (Camp et al., 2001). The second method identifies which instruction leads to the highest test performance combined with the lowest mental effort during training (Kalyuga, Chandler, & Sweller, 1999). And the third method investigates which instruction leads to the highest test performance combined with the lowest mental effort during the test (Paas & Van Merriënboer, 1993, 1994).

Whereas the Camp et al. (2001) study used the first efficiency method, the current study uses the 3D version including test performance, mental effort on training and training time, to determine the efficiency of the training conditions. The reason why we use test performance instead of training performance is that we define training efficiency as not only leading to an optimal learning environment but also to increased ability to flexibly apply skills in new situations (e.g., on the transfer test).

Another important difference between the Camp et al. study (2001) and the current study concerns the amount of transfer tasks. In the current study, the amount of transfer tasks is increased which leads to a larger variation in complexity in these tasks. Overall, the transfer tasks do not only cover variations of the training tasks, but may also be structurally different from the training tasks.

The present study investigates the effects on training efficiency of (1) dynamic task selection vs. non-dynamic task selection with a fixed task sequence, and (2) the use of different learner variables for dynamic task selection, that is, performance, mental effort, and mental efficiency. It is hypothesized that dynamic task selection leads to more efficient training and better transfer performance than non-dynamic task selection (i.e., fixed task sequence). The second hypothesis states that dynamic task selection based on

mental efficiency leads to more efficient training and higher transfer than selection based on performance or mental effort alone. These two hypotheses are used to execute planned comparisons.

Method

Participants

Ninety-one higher education students (M = 20.5 years, SD = 2.29), which were novices in the domain of ATC, participated in this study. The students were randomly assigned to the fixed and experimental conditions in such a way that the performance condition contained 22 participants and the fixed, mental effort, and mental efficiency conditions contained 23 participants each. Men (n = 63) and women (n = 28) were equally distributed across conditions. Since the fixed condition was used as the baseline for defining the scores of the other experimental conditions, the data of this condition were collected first. All participants were in good health and had normal or corrected-to-normal vision. They received € 20 (approximately $ 26) for their participation.

Materials

The ATC-trainer

The ATC-simulator was adapted from training software programmed in Multimedia Toolbook 4.0 and was integrated into the Delphi-interface. Furthermore, a PowerPoint presentation contained an introduction to the ATC domain.

The training software was run on an IBM-compatible PC (Pentium III, 450 MHz) using an IBM 17-inch SVGA monitor (107-MB). In the training, the participants were confronted with simulated dynamic ATC-situations on a radar screen, in which a number of possible conflicts had been built in. In each training task, participants were required to guide moving aircraft to a specific goal position at a specific altitude. While doing this, they had to ensure that all aircraft stayed within controlled airspace and that they kept a minimum vertical and horizontal separation from the other aircraft. Participants were able to change the altitude and the flight direction of all the aircraft in the simulation by typing the desired values into a command table. Their performance was scored on four variables: (a) the time during which any aircraft was flying outside the controlled airspace (time outside airway); (b) the time during which two or more aircraft were flying too close to each other (no separation); (c) the given number of commands, and (d) the number of aircraft that successfully reached their target (gate hits). The interface provided continuously updated

information on these four variables to the participants. An example of a learning task in the ATC training program is depicted in Figure 1.

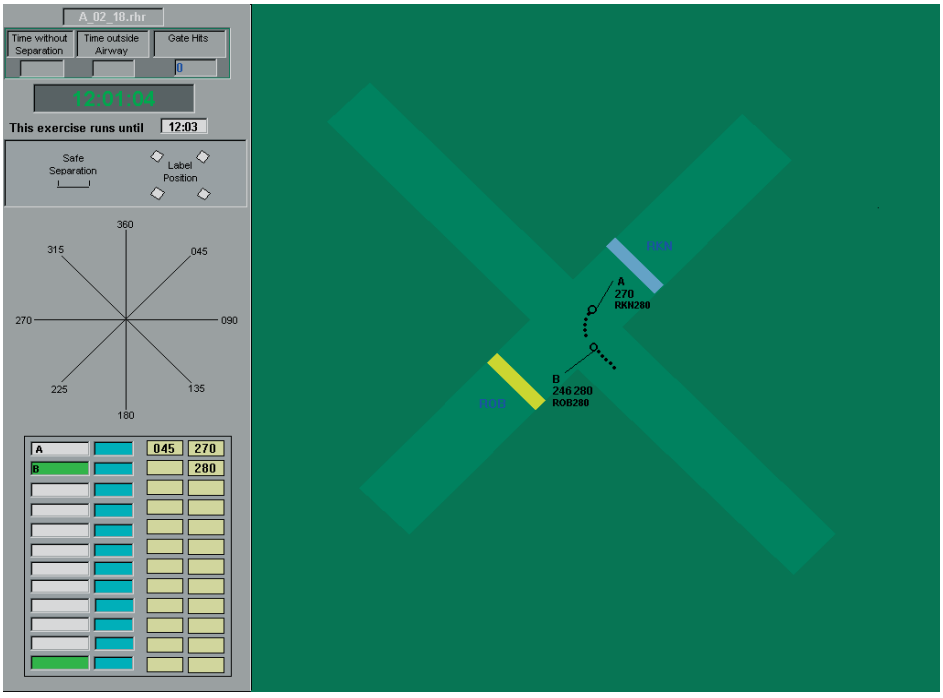

Figure 1. Example of an Air Traffic Control task as used in the training program.

Task complexity

Since the participants of this study are novices in the ATC domain, the overall complexity of the learning tasks was adjusted accordingly prior to training, to enable the participants to perform the tasks. Also, the learning tasks were divided in ten complexity levels which specified the complexity of the tasks based on the number of possible conflicts that was embedded in the task. Four different kinds of conflict were used in this task. The first possible source of conflict (c1) was that a plane's initial flight level differed from its exit flight level. A conflict arose if the flight level of the plane was not changed. The second source of conflict (c2) was that two planes were approaching each other at the same flight level. Again, a conflict arose if the flight level or heading of one of the planes was not changed. The third source of conflict (c3) referred to a situation in which an airplane would have left the airspace (which is forbidden) if no commands were given to change its heading. The fourth possible source of conflict (c4) was somewhat more complex. It referred to the possibility of a conflict due to a command that would normally, in isolation, be beneficial for problem solving but indirectly leads to another conflict. For example, an aircraft

could be given the command to climb to its exit flight level, but this climb could cause a conflict with another aircraft approaching from a different direction.

The different forms of conflict in a task were added to determine the overall complexity of the task. Task complexity was calculated with the following formula: Complexity = c1 + c2 + c3 + 2 (c4). C4 was given double weight, because this type of conflict was more important as it requires the trainee to oversee the whole situation and predict the consequences of his actions. All sources of conflict were scored ordinally for each learning task in all ten complexity levels. For example, a training task of complexity level 5 could contain three conflicts of c2 and one conflict of c4.

Task selection software

In addition to the ATC trainer, the training software used in the experiment also included a program for task selection, the intelligent part of the software. The software controlling the task selection was programmed in Delphi 5.0.

When a student finished a learning task, the mental effort and performance measures were used in the task selection algorithm to calculate mental effort, performance and mental efficiency. The factor 'Method of Task Selection' was used as the independent variable. This factor has four levels: task selection based on (a) a fixed, simple-to-complex sequence of task presentation, (b) mental effort invested in the previous task, (c) performance on the previous task and (d) mental efficiency of the previous task. Depending on the condition the participant was in, the complexity of the next learning task was determined using measures of mental effort, performance, or mental efficiency. Then, a suitable learning task was selected from a database of 77 tasks of complexity levels varying from 1 (e.g., 1 command has to be given to 1 aircraft) to 10 (e.g., 9 commands have to be given to 6 aircraft). This task was then presented to the learner in the training interface. For the participants in the fixed condition this process repeated itself until 20 training tasks, two randomly chosen tasks of every subsequent complexity level, were completed. In the dynamic conditions, three possible outcomes lead to the completion of the training. The first outcome lets participants proceed to the transfer tasks after being presented with 20 training tasks. The second outcome states that a participant has completed the training when s/he has achieved a score that meets the preset performance criteria on two training tasks of the highest complexity. In the last possible outcome, participants complete the training when they have executed all five available training tasks of the highest complexity level. Note that in these last two cases a participant can complete the training while having been presented with less than 20 training tasks.

Selection algorithm

Task selection occurred differently in the four experimental conditions. In the fixed condition, there was no dynamic task selection. Participants in this condition received a total of 20 training tasks, which included two randomly chosen tasks of every complexity level (1, 1, 2, 2, 3, 3, etc.). The scores on the performance variables of the fixed condition were used as a baseline for the scores of the other experimental conditions. As a result, the scores of the participants in the other conditions on these variables were always relative to the scores of the participants in the fixed condition. These data were used to formulate scoring tables from which a mean performance score can be derived.

For all learning tasks in the complexity levels, scales were developed for all performance variables. A file for each complexity level was composed that included all the performance scoring tables for each task that belonged to that specific complexity level. An example of the performance scoring tables for a specific task is depicted in Table 1. The scoring table of a task of complexity level 4 is shown. To obtain the maximum score (100%) for each performance variable of this task, a participant should give four commands and attain four gate hits while no time outside airway occurs and sufficient separation is maintained.

Table 1 *Example of a scoring table for the performance variables.*

	Score				
	0%	25%	50%	75%	100%
N commands	>12	12	9.33	6.67	4
Out of airway	>96	96	64	32	0
No separation	>34	34	22.67	11.33	0
Gate hits	0	1	2	3	4

Table 2 *Mean performance scores.*

Mean performance score		
1	<	31.25%
2	≥	31.25%
3	≥	43.75%
4	≥	56.25%
5	≥	68.75%

Table 3 *Selection table indicating jump size in complexity between training tasks.*

Mental effort	Performance				
	1	2	3	4	5
1	0	+1	+2	+2	+2
2	-1	0	+1	+2	+2
3	-2	-1	0	+1	+2
4	-2	-2	-1	0	+1
5	-2	-2	-2	-1	0

After the relative scores have been appointed for each performance variable, these are added and divided by the number of performance variables. Then, the mean score is divided into five categories (see Table 2); corresponding to an equal division of the performance scores of the fixed condition (i.e., using the 20th, 40th, 60th and 80th percentiles as thresholds). Note that although the number of performance variables usually is four, in several learning tasks of complexity levels 1 and 2 only one aircraft has to be directed to its landing spot. In these tasks the variable 'no separation' does not play a role and therefore this variable is not taken into account to construct the mean performance score.

Dependent on the experimental condition the participant is in, the progression through the training is either based solely on the mental effort score, only on the mean performance score, or on the combination of both scores. The rules specifying the jump sizes can be found in the selection table, which is shown in Table 3. The mean performance scores (1-5) in this table correspond with the scores in Table 2.

In the mental effort condition, task complexity depended on the mental effort learners indicated after finishing a task. The students had to indicate their invested mental effort on a five-point subjective rating scale, with values 1 (very low), 2 (low), 3 (not low, not high), 4 (high), and 5 (very high; see Paas, 1992). The indicated mental effort scores were used directly as the mental effort score of the selection table. For example, if a participant indicated a mental effort of 4 on a task of complexity level 6, then this score is filled in the selection table. As performance is not taken into account in this condition, the performance score is preset on 3 in the selection table. When looking up a mental effort score of 4, it can be determined that the next learning task should be of one complexity level lower than the previous task. In general, a high mental effort leads to easier tasks while low mental effort leads to more difficult tasks.

In the performance condition, both safety and expedition were considered in determining the performance of the specific learning task. The variables that were used to determine performance are time outside airway, time without separation, the number of commands given, and the number of aircraft that successfully reached their target. After the completion of a task, the data of the participants were categorized into the relative scores of the scoring tables of the four performance variables. The mean score was divided into five categories which are the mean performance scores that can be found in the selection table (Table 3). As mental effort is not taken into account in this condition, the mental effort score is preset on 3 in the selection table. When looking up a mean performance score of 5, the next learning task should be of two complexity levels higher than the last presented task. Overall, a high performance leads to more difficult tasks while low performance leads to easier tasks.

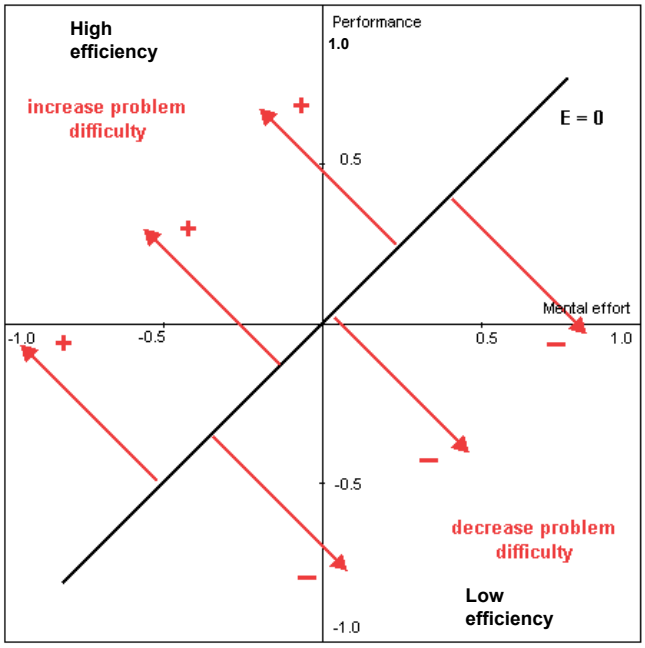

Figure 2. Representation of the effect of mental efficiency on the selection of the complexity of the next learning task.

Filling in both performance and mental effort scores in the selection table determined mental efficiency. When the efficiency score is larger than zero, task complexity is increased. If the efficiency score is smaller than zero, task complexity is decreased (see Figure 2). The reason for this is straightforward. If mental efficiency is larger than zero, the mental effort score is lower than the performance score, indicating that the task was relatively easy. The learner performed relatively well, but invested less mental effort in the learning task than could be expected from his or her performance score. If mental efficiency is smaller than zero, the mental effort score is higher than the performance score, indicating that the task was relatively hard. The learner invested relatively much mental effort in the task, but did not perform accordingly. Task complexity was adjusted on the basis of this argumentation. The exact relation between mental efficiency and change in task complexity can be seen in the selection table, Table 3. For instance, if a participant had a mental effort score of 4 and a mean performance score of 5, then task complexity was increased with one level (+1).

Transfer test

After the training, the participants were required to solve ten transfer tasks. Half of these tasks were structurally similar to the training tasks, but the aircraft

had different values. The other half of the tasks was structurally different from the training tasks in several ways. First, new aircraft frequently appeared in the interface. Second, some of the aircraft had a different speed than other aircraft while all aircraft had the same speed in the training tasks. Finally, the number of aircraft that had to be directed to their appropriate landing spot was larger than in any of the training tasks.

Procedure

The participants received a condition-specific training program on ATC. They were unaware of the conditions of the experiment, and therefore did not know how their training tasks were selected. First, all participants were given an introduction to the field of ATC within the training software. In this introduction, the knowledge that was required for the training was presented and the participants were shown how to give commands to the aircraft and were familiarized with the way the aircraft react to the commands (delay etc.). Participants were able to return to this introduction at any time during the training, and were explicitly advised to do so if they had any doubts about their understanding of the learning tasks. After the introduction, the participants had to complete a practice task in which they had to practice giving commands (direction and altitude) to one aircraft. After the completion of this practice task, the participants proceeded with the actual ATC-training. All participants could continue with a next learning task when they had completed the previous task, meaning that differences in training time could occur in all conditions. Depending on the amount of possible conflict in the task, task complexity varied between 1 and 10, with 10 being the most difficult type of task. All participants started with a task of complexity level 1. After their first task, depending on the experimental condition they were in, the next learning task was selected on the basis of their performance, experienced mental effort, or a combination of both measures. The duration of the whole experiment varied from 1.5 to 2 hours, in which the training tasks were solved, followed by 10 transfer tasks.

Results

First, the results of five dependent variables will be given of the four experimental conditions in order to gain insight in the task selection process. These variables are: number of learning tasks that was completed before reaching the highest complexity level, highest complexity level that was reached during training, size of the jumps between complexity levels, total number of training tasks, and training time. Then, the results for performance and mental effort during training are provided. Finally, the results on the dependent variables performance and mental effort on the transfer test are

Table 4 *Overview of results.*

	Method of Task Selection							
	Fixed		Mental effort		Performance		Mental efficiency	
Dependent variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Training					
N tasks before highest complexity	18.0	0	15.0	1.4	7.1	3.5	7.8	3.4
Highest complexity	10.0	0	6.7	2.6	9.8	1.1	9.5	1.4
Jump size	0.45	0	0.42	0.10	0.65	0.08	0.64	0.11
Total N of learning tasks	20.0	0	19.9	0.4	10.1	3.9	14.5	3.8
Training time	3154.1	161.0	2763.6	390.0	1623.1	642.8	2354.0	564.7
Performance	76.2	13.1	81.1	11.4	75.5	11.0	71.0	9.6
Mental effort	3.2	0.5	2.6	0.3	3.0	0.4	2.9	0.4
			Test					
Performance	68.5	15.6	70.5	10.2	69.3	13.3	67.8	14.0
Mental effort	3.4	0.5	3.5	0.6	3.4	0.5	3.1	0.5
Training efficiency	-0.88	1.23	0.25	0.56	0.56	1.30	0.09	1.16

given, as well as the results of the training efficiency. Analyses of variance (ANOVAs) and planned comparisons were used to analyze the data.

Training Phase

Training effects

The four conditions of the independent variable Method of task selection were compared on five dependent variables to assess the differences in the training phase. The values on the first four variables in the fixed condition were not included in the analysis of variance because they were preset on constant numbers. In other words, there was no variation between the participants in the fixed condition. On the variable training time however, variation was possible because all participants could continue with a next task as soon as s/he had completed the previous task. Means and standard deviations are provided in Table 4.

With regard to the number of learning tasks that was completed before reaching the highest complexity level, an ANOVA revealed a significant effect for the factor Method of Task Selection, $F(2, 44) = 14.1$, $MSE = 10.8$, $p < .0001$. Planned comparisons showed that the mental efficiency condition ($M = 7.8$, $SD = 3.4$) practiced less learning tasks before reaching the highest complexity level ($t(44) = -3.12$, $p < .01$) than the mean number of learning tasks of the performance condition and the mental effort condition ($M = 11.1$, $SD = 2.45$; with $M = 7.1$, $SD = 3.5$ for the performance condition and $M = 15.0$, $SD = 1.4$ for the mental effort condition).

With regard to the highest complexity level that was reached during the training, a significant effect was found for Method of Task Selection, $F(2, 65) = 20.5$, $MSE = 3.31$, $p < .0001$. Planned comparisons showed that the mental efficiency condition ($M = 9.5$, $SD = 1.4$) reached a higher complexity level ($t(65) = 2.72$, $p < .01$) than the mean highest reached complexity level of the performance condition and the mental effort condition ($M = 8.3$, $SD = 1.85$; with $M = 9.8$, $SD = 1.1$ for the performance condition and $M = 6.7$, $SD = 2.6$ for the mental effort condition).

The absolute jump size in complexity level depended on the learners' mental effort, performance, or mental efficiency on the previous task. Jumps in complexity level in the dynamic conditions could be both negative and positive, leading to easier or more difficult tasks, respectively. In the analysis, the focus was on differences in absolute jump size. Using ANOVA, a main effect of Method of Task Selection was found, $F(2, 65) = 28.6$, $MSE = .01$, $p < .0001$. Planned comparisons showed that the mental efficiency condition ($M = .64$, $SD = .11$) made larger jumps ($t(65) = 3.43$, $p < .01$) than the mean jump size score of

the performance condition and the mental effort condition (M = .54, SD = .09; with M = .65, SD = .08 for the performance condition and M = .42, SD = .10 for the mental effort condition).

An ANOVA of the total number of learning tasks revealed a main effect of Method of Task Selection, $F(2, 65) = 55.8$, $MSE = 9.78$, $p < .0001$. Planned comparisons showed that the mental efficiency condition (M = 14.5, SD = 3.8) did not practice less tasks ($t(65) = -.65$, $p = .52$) than the mean number of tasks of the performance condition and mental effort condition (M = 15.0, SD = 2.15; with M = 10.1, SD = 3.9 for the performance condition and M = 19.9, SD = .4 for the mental effort condition).

With regard to training time, an ANOVA revealed a significant effect for the factor Method of Task Selection, $F(3, 87) = 42.6$, $MSE = 225376.6$, $p < .0001$. Planned comparisons showed that the fixed condition (M = 3154.1, SD = 161.0) needed more time to complete the training ($t(87) = 7.92$, $p < .0001$) than the mean training time of the performance condition, the mental effort condition, and the mental efficiency condition (M = 2246.9, SD = 532.5; with M = 1623.1, SD = 642.8 for the performance condition, M = 2763.6, SD = 390.0 for the mental effort condition, and M = 2354.0, SD = 564.7 for the mental efficiency condition). No difference was found between the mental efficiency condition ($t(87) = 1.32$, $p = .190$) and the mean training time of the performance and mental effort conditions (M = 2193.4, SD = 516.4).

Performance and mental effort. Main effects of Method of Task Selection were found on the variables used to determine mental efficiency; $F(3, 87) = 3.1$, $MSE = 129.1$, $p < .05$ for performance, and $F(3, 87) = 8.3$, $MSE = .17$, $p < .0001$ for mental effort. Planned comparisons revealed no difference ($t(87) = .11$, $p = .92$) in performance between the fixed condition (M = 76.2, SD = 13.1) and the mean performance score of the performance condition, the mental effort condition, and the mental efficiency condition (M = 75.9, SD = 10.67; with M = 75.5, SD = 11.0 for the performance condition, M = 81.1, SD = 11.4 for the mental effort condition, and M = 71.0, SD = 9.6 for the mental efficiency condition). However, planned comparisons did show that the mental efficiency attained a lower training performance ($t(87) = -2.51$, $p < .05$) than the mean performance of the performance and mental effort conditions (M = 78.3, SD = 11.2).

Furthermore, planned comparisons showed that the fixed condition (M = 3.2, SD = .5) invested more mental effort during training than the mean score of the performance condition, the mental effort condition, and the mental efficiency condition (M = 2.8, SD = .38; with M = 3.0, SD = .4 for the performance condition, M= 2.6, SD = .3 for the mental effort condition, and M = 2.9, SD = .4 for the mental efficiency condition). No difference was found between the efficiency condition and the mean score of the performance and mental effort condition (M = 2.8, SD = .37).

Transfer Test Phase

Performance and Mental Effort

When analyzing the data on the transfer test, no differences were found between the different methods of task selection. There were no significant differences in performance or mental effort between the four experimental groups on the ten transfer tasks (all *F*s < 1).

Training efficiency

The training efficiency was determined using the following formula:

$$E = \frac{P - ME - TT}{\sqrt{3}}$$

In this formula, E = mental efficiency, ME = mental effort during training, P = test performance, and TT = total training time. Using an ANOVA, a significant effect was found for Method of Task Selection, $F(3, 87) = 7.3$, $MSE = 1.21$, $p < .0001$. Planned comparisons showed that the fixed condition ($M = -.88$, $SD = 1.23$) was less efficient ($t(87) = -4.46$, $p < .0001$) than the mean efficiency score of the performance condition, mental effort condition, and mental efficiency condition ($M = .3$, $SD = 1.02$; with $M = .56$, $SD = 1.30$ for the performance condition, $M = .25$, $SD = .56$ for the mental effort condition, and $M = .09$, $SD = 1.16$ for the mental efficiency condition). No difference ($t(87) = -1.11$, $p = .27$) was found between the mental efficiency and the mean efficiency score of the performance and mental effort conditions ($M = .41$, $SD = .93$). The means and standard deviations are provided in Table 4.

Discussion

The main hypothesis of this study that dynamic task selection leads to more efficient training than a fixed task sequence was confirmed. The results show that the training efficiency of the conditions in which learning tasks were dynamically selected was significantly higher than the efficiency of the fixed condition. The specific hypothesis, that dynamic task selection based on mental efficiency would lead to more efficient training and better transfer than selection based on performance or mental effort alone, was not confirmed.

The significant efficiency effect shows that dynamic task selection leads to more efficient training than a fixed, predetermined training sequence which does not adjust to the individual student. Although the fixed condition did attain the same performance score as the three dynamic conditions, its costs in terms of time and mental effort to achieve this performance level were substantially higher. In line with the prediction based on cognitive load theory, adjusting the training tasks to the learners' cognitive state in the dynamic

conditions was more efficient than without this adjustment in the fixed condition. Although no support was found for the hypothesis that the mental efficiency condition would lead to more efficient training than the other two dynamic conditions, the mental efficiency condition appeared to be somewhat more effective during training than the mental effort and performance conditions. The participants in the mental efficiency condition needed less learning tasks to reach the highest complexity level, reached a higher overall complexity level, and made larger jumps than the students in the mental effort and performance conditions. However, while no differences were found in terms of training time and amount of invested mental effort, the mental efficiency condition did attain a lower performance score than the mental effort and performance conditions. An explanation for this could be that since they made larger jumps and reached the highest complexity level faster, they practiced more complex learning tasks than the other dynamic conditions. Although the participants in the mental efficiency condition practiced more complex tasks, they did not invest more mental effort than the participants in the other dynamic conditions. Overall, the task selection method of the mental efficiency condition was not more efficient than the task selection methods of the two other dynamic conditions.

Besides the difference between the fixed condition and the dynamic conditions in training efficiency, no differences were found between the dynamic conditions in performance and invested mental effort on the transfer tasks. Several possible explanations can be given for this. First, the transfer tasks could have been too difficult for all participants to solve. If this is true, a floor effect should be present indicating low performance scores and high invested mental effort. A closer look at the overall performance of all conditions (69%) and the amount of invested mental effort (3.4), showed that floor effects can be excluded. However, since the performance means and standard deviations of all the conditions are in the same range, a ceiling effect might have occurred. Though the performance means are relatively high, it seems that participants were unable to attain the highest possible score of 100%.

Another possibility, which is tentative but interesting as well, was presented by Camp et al. (2001). Complex cognitive skills consist of a number of sub-skills, which can be either recurrent or non-recurrent (van Merriënboer, 1997). Recurrent skills are rule-based skills that are learned through the process of rule automation, which involves a vast amount of practice on the same task (e.g., Anderson, 1987). Non-recurrent skills are knowledge-based skills that are learned through the process of schema construction, which is stimulated by receiving a varied sequence of tasks (e.g., Singley & Anderson, 1989). These two different skills can both improve transfer of training. According to some authors (e.g., van Merriënboer, 1997), transfer can occur because the learning tasks and the transfer tasks share identical elements: familiar aspects of a task

are performed rule-based, because of the availability of domain-specific rules. But, at the same time, unfamiliar aspects of a task are performed with the help of cognitive schemata. The new task is reorganized in such a way that it can be understood in terms of these cognitive schemata.

Camp et al. (2001) proposed that the performance condition in their study, which resulted in much practice on tasks with the same complexity, might have fostered rule automation. On the other hand, the mental effort and mental efficiency conditions, in which practice was more variable regarding the complexity, might have fostered the construction of cognitive schemata. In the current study, the task selection algorithm of the mental effort condition led to low-variable practice and, probably, stimulated rule automation. In contrast, the performance and mental efficiency conditions resulted in more variable practice and can be expected to have stimulated the process of schema construction. Despite these different effects between the study of Camp et al. and this study, it is important to note that a task selection algorithm can be constructed in such way that it can influence different cognitive processes. It would be interesting to further investigate this notion and use performance tests that are sensitive to the different cognitive learning processes of rule automation and schema construction.

The proposed extended efficiency formula revealed a difference between the fixed condition and the experimental conditions. The inclusion of training time in the formula is a refinement which enables the formula to take other differences into account besides the performance and mental effort differences. One could also use time-on-task as the third dimension of the efficiency formula in case the tasks are flexible in time to complete. However, the variation in maximum time in which a task had to be completed was small in the current study. Therefore, we chose to use total training time instead of time-on-task as the third dimension of the 3D efficiency formula. For future research, it could be interesting to use time-on-task as the third dimension of the 3D efficiency formula as well as using time on task as another determinant for task selection during training.

Furthermore, it is important to develop a transfer test that is highly sensitive to the differential effects of the experimental conditions. A reaction time test could be a useful technique because learners have to decide as fast and as accurately as possible whether a specific ATC situation includes a conflict. For example, by presenting trainees a screen dump of an ATC situation for a few seconds, one could test the amount of automation and elaboration of knowledge. After presenting such a situation shortly, a multiple choice question can be given from which the trainee should choose the correct order of aircraft to which commands have to be given in order to ensure that all aircraft land safely. For example, when 3 aircraft (A, B, C) are presented in the screen dump, the trainee should pick the right order in giving commands to the aircraft (e.g.,

B, C, A). To test more extensive elaboration of knowledge, one could also provide a multiple choice question which not only consists of the order of commands but also the exact commands. For example, when a screen dump has 2 aircraft (A, B), the right solution could be (B: 240 altitude, 90 right turn, A: 220 altitude, 315 left turn).

In conclusion, the results regarding the first hypothesis of this study supported the idea that adapting training to the individual needs of the student makes training more efficient. No evidence was found to support the second hypothesis, which stated that task selection based on mental efficiency would lead to more efficient training and higher transfer than selection based on either performance or mental effort alone. However, the mental efficiency condition did show several training benefits over the other dynamic conditions. The use of the extended 3D efficiency formula was proven to be successful in differentiating between the fixed condition and the dynamic conditions. This result is encouraging for further experimentation and refinement of the formula with regard to increasing the differentiation capability between the dynamic conditions. Furthermore, more research is needed to develop a transfer test that is sensitive enough to measure the possible effects due to the different training effects of the experimental methods.

The results of this study, combined with the results of Camp et al. (2001), indicate that for training with ITS, using dynamic task selection can be worthwhile for the reduction of both training time and costs. This can be particularly interesting for domains in which training time and costs are of great importance, like aviation and industry. The potential of mental effort and mental efficiency as variables to be used in dynamic task selection needs further study.

References

Anderson, J. R. (1987). Skill acquisition: compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Camp, G., Paas, F., Rikers, R. M. J. P., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

EATMP Human Resources Team (2001). *Guidelines for Selection Procedures and Tests for Ab Initio Trainee Controllers (Revised)*, HRS/MSP-002-GUI-01. Brussels, Belgium: EUROCONTROL.

Galster, S. M., Duley, J. A., Masalonis, A. J., & Parasuraman, R. (2001). Air Traffic Controller performance and workload under mature free flight: conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology, 11*, 71-93.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Kashihara, A., Hirashima, T., & Toyoda, J. (1995). A cognitive load application in tutoring. *Journal of User Modeling and User-Adapted Instruction, 4*, 279-303.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: a cognitive-load approach. *Journal of Educational Psychology, 84*, 429-434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Tuovinen, J. E., Tabbers, H. K., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Parasuraman, R., Molly, R., & Singh, I. L. (1993). Performance consequences of automation induced complacency. *International Journal of Aviation Psychology, 3*, 1-23.

RTCA (1995). *Reports of the RTCA board of director's select committee on free flight*. Unpublished manuscript.

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

Singley, M. K., Anderson, J. R. (Eds.) (1989). *The transfer of cognitive skills*. Cambridge, MA: Harvard University Press.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.

Sweller, J. (1999). *Instructional Design in Technical Areas*. Melbourne, Australia: ACER Press.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*, 133-152.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model*. Englewood Cliffs, NJ: Educational Technology Publications.

Chapter 4 – Personalized Task Selection in Air Traffic Control: Effects on Training Efficiency and Transfer [1]

Abstract

The differential effects of four task selection methods on training efficiency and transfer in a computer-based training for Air Traffic Control were investigated. Two personalized conditions were compared with two corresponding yoked control conditions. The hypothesis that personalized dynamic task selection leads to more efficient training than non-dynamic task selection was partially confirmed. However, the hypothesis that dynamic task selection based on personalized efficiency leads to more efficient training than dynamic task selection based on learner control (i.e., personalized preference) was not supported. The results are discussed and suggestions are given for future research.

Introduction

Developments in Training Methods for Complex Cognitive Skills

Within the aviation domain there is a serious shortage of well-trained Air Traffic Controllers, mainly due to the yearly increasing crowdedness of the airspace (Galster, Duley, Masalonis, & Parasuraman, 2001). One possible solution to this problem is increasing the efficiency of Air Traffic Control (ATC) training. Efficient training offers trainees an optimal learning environment in which they can acquire skills quickly and adequately while they also learn how to apply these skills flexibly to new situations (i.e., transfer).

A theoretical comparison of training methods for complex cognitive skills (*Chapter 2:* Salden, Paas, & van Merriënboer, in press) showed that these methods have evolved in three important ways during the last three decades. The first change marks a shift in focus from non-dynamic to dynamic methods. Although both methods take prior knowledge into account in the development of the training, non-dynamic approaches can only determine the sequence of learning tasks prior to the start of the training. Yet dynamic approaches also have the possibility to make adjustments in the task sequence during training. The second change reflects the development of training methods from being part-task based to whole-task based. Part-task methods might be useful for a complex task (e.g., learning to drive a car) where the trainee is not able to practice the task in its entirety at the start of the training. However, whole-task methods are more appropriate when parts of a task are strongly interrelated, which makes it very difficult to define and train meaningful parts without

[1] *This chapter is based on: Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (2004). Personalized task selection in Air Traffic Control: Effects on training efficiency and transfer. Manuscript submitted for publication.*

compromising sophisticated understanding. The third change shows a shift from group-based to personalized methods. Whereas group-based training can be very useful in terms of allocated time and resources, the intricate nature of complex cognitive skills imposes different demands on each individual student.

The use of a personalized and dynamic whole-task method is believed to be strongly related to increased training efficiency (*Chapter 3:* Salden, Paas, Broers, & van Merriënboer, 2004). The group-based non-dynamic training methods, currently being used in ATC training programs, present students with a preset order and complexity of learning tasks, and do not have the adaptive ('dynamic') ability to make adjustments in complexity and task order during training. Reports on current ATC training methods (e.g., EATMP Human Resources Team, 2001) show that these non-dynamic methods exhibit a high dropout rate of ATC students.

System-Controlled Task Selection vs. Learner-Controlled Task Selection

As the role of the computer increased significantly, training programs became more and more computer-based, enabling trainers to make adjustments in the order and complexity of the learning tasks during the training phase. Many Intelligent Tutoring Systems (ITS) use a student model in order to keep track of the individual trainee's history of the tasks and the corresponding performance. A student model builds a knowledge base of the trainee, and updates that knowledge base as the trainee progresses through the learning tasks. The progress of the trainee is checked on the basis of comparing the trainee's performance to the learning objectives that were specified prior to training. After this comparison, the system-controlled selection rules indicate the appropriate next learning task to present to the learner.

Besides such system-controlled task selection, learner-controlled selection may offer another form of personalized dynamic task selection because it gives the students control over what learning tasks they want to practice next. While a clear definition of learner control is missing (Reeves, 1993), most studies in the field of computer-based training operationalize it in two ways: Either learners are given the option to request additional instructional material or they are given the option to bypass instructional material (Crooks & Klein, 1996). The basic theoretical claim for potential positive effects of learner control (i.e., personalized preference) is that trainees are able to select the appropriate tasks to practice while avoiding a possible overload of their cognitive system, thereby increasing the effectiveness and efficiency of learning (e.g., Borsook & Higginbotham-Wheat, 1991). However, several studies show that low-ability learners experience problems with the control they are given (e.g., Bell & Kozlowski, 2002; Niemic, Sikorski, & Walberg, 1996; Steinberg, 1977, 1989; Williams, 1993). A possible explanation is that the given level of control is often not compatible with the learners' abilities.

According to Bell and Kozlowski (2002), it is critical to design instructional material that provides learners with a level of control they are able to handle. Furthermore, the 'expertise reversal effect' (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 1998, 2001) indicates that the trainees' increasing expertise level is probably the most important determinant for deciding on the appropriate level of freedom that is given to them. For example, van Merriënboer, Schuurman, de Croock, and Paas (2002) found that learners who are given the possibility to choose the task format in the domain of computer programming are well able to select their own learning tasks.

Measures for System-Controlled Task Selection

Research in the context of cognitive load theory (for an overview see Paas, Renkl, & Sweller, 2003) has shown that cognitive load is a crucial factor in the training of complex cognitive skills. Although cognitive load is sometimes measured (e.g., Kashira, Hirashima, & Toyoda, 1995), usually only performance measures such as speed and accuracy are used as determinants for personalized task selection, for instance in Intelligent Tutoring Systems. In order to obtain a good indication of the cognitive load that is imposed on a person's cognitive system, mental effort measurements such as subjective rating scales are used. While individual measures of performance and mental effort can be used as indicators of the cognitive demands a certain task places on the learner, the combination of both measures is considered a superior estimate of these demands in the dynamic selection of learning tasks (Paas & van Merriënboer, 1993).

When trainees achieve the same performance scores, mental effort ratings might be able to reveal differences that would remain otherwise unnoticed. For example, it is quite feasible for two people to attain the same performance levels. However, while one of them experiences a very high cognitive load and needs to work laboriously through a very effortful process, the other person may experience a low cognitive load and may reach the same performance level with a minimum of effort. Most people would agree that the next learning task should be less difficult for the first person than for the second person. Since the combination of performance and mental effort measures provides a clear picture of the state of the student's cognitive system at a certain moment in training, we claim that personalized task selection, and consequently training efficiency can be improved by taking these measures into account. When using both measures for personalized dynamic task selection, the selected learning tasks are believed to be better adjusted to the student's cognitive schemata and cognitive capacity, hence leading to high training efficiency.

A first indication for the claim that the use of a combined measure of performance and mental effort scores (i.e., personalized efficiency) leads to

more efficient personalized training was found in two recent studies of Camp, Paas, Rikers, and van Merriënboer (2001), and Salden et al. (*Chapter 3:* 2004). Both studies compared four methods of task selection in the ATC domain. The first method presented tasks in a fixed, predetermined sequence from simple to complex. The other three methods presented the tasks dynamically, based on either performance or mental effort, or the combination of both (i.e., mental efficiency). Results showed that personalized dynamic task selection leads to more efficient training than group-based non-dynamic task selection. Although personalized dynamic task selection based on mental efficiency did not lead to more efficient training and better transfer than personalized dynamic task selection based on performance or mental effort alone, it revealed show several training benefits.

The combined measure of performance and mental effort scores, as used in the aforementioned studies (Camp et al., 2001; *Chapter 3:* Salden et al., 2004), has been proposed as a measure of mental efficiency by Paas and van Merriënboer (1993; see also Paas, Tuovinen, Tabbers, & Van Gerven, 2003). These authors present a calculational approach for combining measures of mental workload and task performance that allows one to obtain information on the relative efficiency of instructional conditions. Based on Ahern and Beatty's (1979) efficiency view on learning, it is proposed that learners' behavior in a certain learning condition is more efficient if (1) their performance is higher than might be expected on the basis of their invested mental effort, and/or (2) their invested mental effort is lower than might be expected on the basis of their performance. Thus, high performance combined with low mental effort is more efficient than low performance combined with high mental effort.

Hypotheses

In this study it is hypothesized that personalized dynamic task selection leads to more efficient training and better transfer performance than non-dynamic task selection. Since both dynamic conditions personalize the task selection either by utilizing mental efficiency or by allowing learners to determine their own training sequence, these conditions are expected to lead to superior training efficiency.

The second hypothesis states that dynamic task selection based on personalized efficiency leads to more efficient training and better transfer than selection based on personalized preference. While the efficiency condition adjusts to the needs of the individual learners, learners in the personalized preference condition have to select training tasks by themselves. The literature on learner control suggests that especially low prior knowledge students, like the ones used in the current study may be overwhelmed by the freedom given to them (e.g., Bell & Kozlowski, 2002; Niemic, Sikorski, & Walberg, 1996).

Method

Participants

Sixty higher education students (M = 20.3 years, SD = 2.35), who were novices in the domain of Air Traffic Control (ATC), participated in this study. The students were randomly assigned to the four experimental conditions in such a way that each condition contained 15 participants, and that men (n = 48) and women (n = 12) were equally distributed across conditions. All participants were in good health and had normal or corrected-to-normal vision. They received € 20 (approximately $ 26) for their participation.

Materials and Procedure

The experimental materials consisted of a learning phase in which learners were presented with an introduction to the domain followed by a training using an ATC software program where they had to work on learning tasks at different levels of complexity. The selection of the learning tasks depended on the experimental condition that learners had been assigned to. After the training, learners were presented with transfer tasks to register their learning outcomes.

Learner control is defined as being the ability of the learner to choose the complexity level of the learning task s/he wants to practice. In the terms of Crooks and Klein (1996), learners are given the option to bypass instructional material. To avoid confusion due to the two control conditions, learner control will be called personalized preference in the remainder of this article.

Introduction

At the beginning of the experiment, the participants were given a Microsoft ® PowerPoint ® presentation containing an introduction to the ATC domain. In this presentation, the knowledge that was required for the training was presented and the participants were shown how to give commands to the aircraft. Participants were able to return to this introduction at any time during the training, and were explicitly advised to do so if they had any doubts about their understanding of the learning tasks. After the presentation, the participants were given a pre-training in which they were presented a total of four learning tasks: two tasks from each of complexity levels 1 and 2. In this pre-training they could practice basic skills such as giving commands in direction or in altitude to one or two aircraft, hereby getting familiarized with the way aircraft react to the commands. After the completion of this pre-training, the participants proceeded with the actual ATC-training.

The ATC-Trainer

The ATC-simulator was an adapted version of training software programmed in Multimedia Toolbook 4.0 and was integrated in a Delphi-interface. The training software was run on an IBM-compatible PC (Pentium III, 450 MHz) using an IBM 17-inch SVGA monitor. In the training, the participants were confronted with simulated dynamic ATC-situations on a radar screen, in which a number of possible conflicts had been built in. Participants were required to guide moving aircraft to a specific goal position at a specific altitude. While performing this task, they had to ensure that all aircraft stayed within the controlled airspace and that a minimum vertical and horizontal separation from the other aircraft was maintained. Participants were able to change the altitude and the flight direction of all the aircraft in the simulation by typing the desired values into a command table. Their performance was scored on four variables: (a) the time during which any aircraft was flying outside the controlled airspace (time outside airway); (b) the time during which two or more aircraft were flying too close to each other (no separation); (c) the given number of commands, and (d) the number of aircraft that successfully reached their target (gate hits). The interface provided continuously updated information on these four variables to the participants. The interface is depicted in Figure 1.

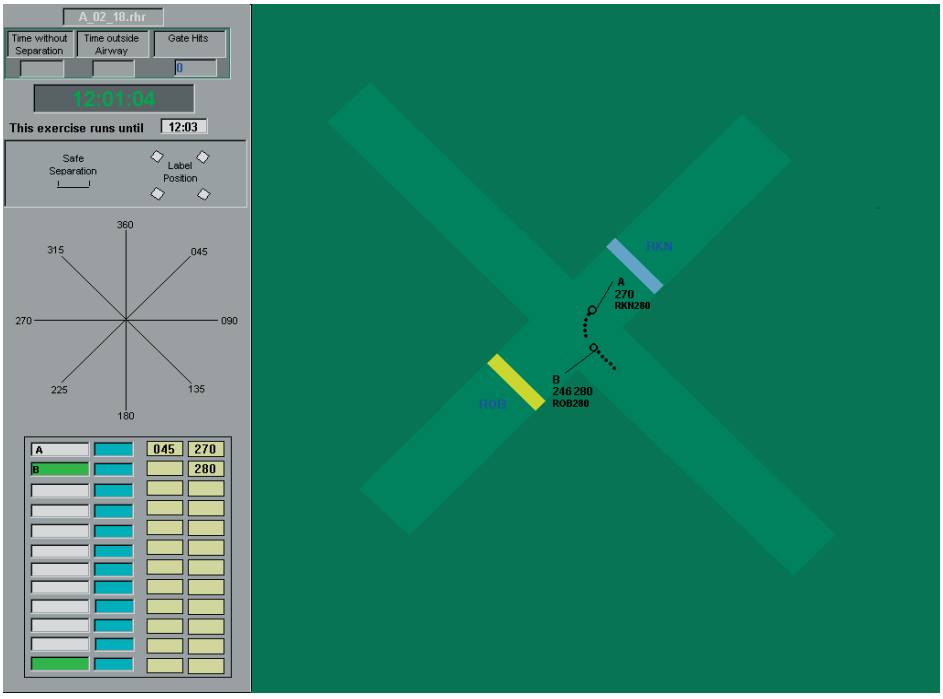

Figure 1. Example of an Air Traffic Control task as used in the training program.

Learning Task Complexity

Prior to training, 77 learning tasks were divided into ten complexity levels varying from 1 to 10. These levels specified the complexity of the tasks based on the number and importance of possible conflicts that was embedded in the task. Four different kinds of conflict could arise during a learning task, either singularly or comprised. The first possible source of conflict (c1) was that a plane's initial flight level differed from its exit flight level. A conflict arose if the flight level of the plane was not changed. The second source of conflict (c2) was that two planes were approaching each other at the same flight level. Again, a conflict arose if the flight level or heading of one of the planes was not changed. The third source of conflict (c3) referred to a situation in which an airplane would have left the airspace (which is forbidden) if no commands were given to change its heading. The fourth possible source of conflict (c4) was somewhat more complex. It referred to the possibility of a conflict due to a command that would normally, in isolation, be beneficial for problem solving but indirectly led to another conflict. For example, an aircraft could be given the command to climb to its exit flight level, but this climb could cause a conflict with another aircraft approaching from a different direction.

The different forms of conflict in a task were added to determine the overall complexity of the task. In consultation with professional Air Traffic Controllers from the Eurocontrol Institute of Air Navigation Services, the task complexity was determined with the following formula: Complexity = c1 + c2 + c3 + 2 (c4). The parameter c4 was given double weight, because this type of conflict was more complex as it required the trainee to oversee the whole situation and predict the consequences of his actions. All sources of conflict were scored ordinally for each learning task in all the complexity levels. For example, a learning task of complexity level 5 could contain three conflicts of c2 and one conflict of c4. Depending on the amount of possible conflicts in the tasks, task complexity varied between 1 and 10, with 10 being the most difficult type of task. Since all participants were presented with learning tasks of complexity levels 1 and 2 in the pre-training, they started the training with a learning task of a complexity level higher than 2.

Overall, three possible outcomes could lead to the completion of the training. The first outcome let participants proceed to the transfer tasks after being presented with 20 learning tasks. This amount of tasks ensured sufficient variation over the ten complexity levels. The second outcome stated that a participant had completed the training when s/he had achieved a score that met the preset performance and mental effort criteria on two learning tasks of the highest complexity. In the last possible outcome, participants completed the training when they have executed all five available learning tasks of the highest complexity level. Note that in these last two cases a participant could complete the training while having been presented with less than 20 learning tasks.

Design and Dependent Variables

Design

Learners were trained accordingly to the factor 'Method of Task Selection' in one of four conditions: (1) personalized efficiency condition, (2) yoked efficiency condition, (3) personalized preference condition, and (4) yoked preference condition.

(1) Personalized efficiency condition. The selection of learning tasks in the personalized efficiency condition was based on the combination of performance and mental effort measures. The task selection table (Table 1) shows that both performance and mental effort were scored on a 5-point scale and the difference between these two factors marks the complexity level for the next learning task.

More specifically, performance was measured on four variables: number of commands, number of gate hits, time outside airway, and time without separation. Mental effort was measured using a 5-point subjective rating scale (1 = very low; 5 = very high) which participants had to fill in after each completed task. These subjective ratings of mental effort were directly used in Table 1 for determining the complexity level of the next learning task.

Since the participant population was similar to a previous experiment (*Chapter 3:* Salden et al., 2004), the scores on the performance variables of this previous experiment's fixed condition were used as a baseline for the scores of the personalized efficiency condition in the current study. In that fixed condition, participants received a total of 20 learning tasks, which included two randomly chosen tasks of every complexity level (1, 1, 2, 2, 3, 3, etc.). The data from the fixed condition of Salden et al.'s (*Chapter 3:* 2004) experiment were used to formulate scoring tables from which a mean performance score could be derived.

Also, scoring scales were developed for all performance variables. For each complexity level a file was composed that included all the performance scoring tables for each task that belonged to that specific complexity level. An example of the performance scoring tables for a specific task of complexity level 4 is depicted in Table 2. To obtain the maximum score (100%) for each performance variable of this task, a participant should give four commands and attain four gate hits while no time outside airway occurs and sufficient separation is maintained.

Table 1 *Selection table indicating jump size in complexity between training tasks.*

Mental effort	Performance				
	1	2	3	4	5
1	0	+1	+2	+3	+4
2	-1	0	+1	+2	+3
3	-2	-1	0	+1	+2
4	-3	-2	-1	0	+1
5	-4	-3	-2	-1	0

Table 2 *Example of a scoring table for the performance variables.*

	Score				
	0%	25%	50%	75%	100%
N commands	>12	12	9.33	6.67	4
Out of airway	>96	96	64	32	0
No separation	>34	34	22.67	11.33	0
Gate hits	0	1	2	3	4

Table 3 *Mean performance scores.*

Mean performance score		
1	<	31.25%
2	≥	31.25%
3	≥	43.75%
4	≥	56.25%
5	≥	68.75%

After the relative scores had been appointed for each performance variable, these were added and divided by the number of performance variables. Then, the mean score was divided into five categories (see Table 3); corresponding to an equal division of the performance scores of the fixed condition, using the 20th, 40th, 60th, and 80th percentiles as thresholds. Note that although the number of performance variables is usually four, in several learning tasks of complexity levels 1 and 2 only one aircraft has to be directed to its landing spot. In these tasks the variable 'no separation' does not play a role and therefore this variable is not taken into account to construct the mean performance score.

Filling in both performance and mental effort scores in the selection table (Table 1) determined mental efficiency. When the efficiency score is larger than zero, task complexity is increased. If the efficiency score is smaller than zero, task complexity is decreased (see Figure 2). The reason for this is straightforward. If mental efficiency is larger than zero, the mental effort score (X axis) is lower than the performance score (Y axis), indicating that the task was relatively easy. The learner performed relatively well and had invested less mental effort in the learning task than could be expected from his or her performance score. If mental efficiency is smaller than zero, the mental effort score is higher than the performance score, indicating that the task was

relatively hard. In this situation, the learner invested relatively much mental effort in the task, but did not perform accordingly. Task complexity was adjusted on the basis of this argumentation. The exact relation between mental efficiency and change in task complexity can be seen in Table 1. The mean performance scores in this table correspond to the scores in Table 3. For instance, if a participant indicated a mental effort score of 2 and a mean performance score of 5, then task complexity was increased with three levels (+3). It can be derived from Table 1 that the participants in this condition had a maximum jump size of 4 (e.g., performance = 5 and mental effort = 1; or performance = 1 and mental effort is 5).

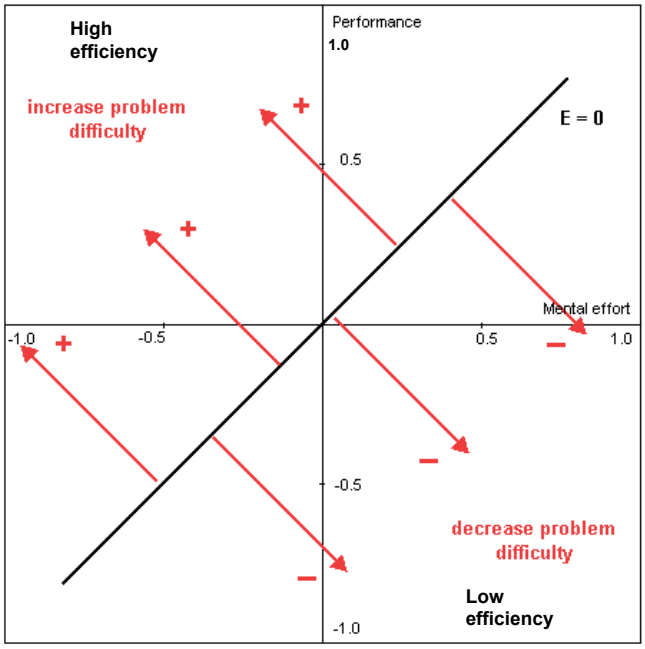

Figure 2. Representation of the effect of mental efficiency on the selection of the complexity of the next learning task.

(2) Yoked efficiency condition. The participants in the yoked efficiency condition were presented with a fixed training sequence. The individual training sequences of the participants in the personalized efficiency condition were divided over the participants in the corresponding yoked condition. Each training sequence was allocated only once to one participant in the yoked efficiency condition.

(3) Personalized preference condition. After completing the pre-training a window popped up in which the participants in the personalized preference condition could choose the complexity level of a learning task that they wanted to practice. Based on their selection of a certain complexity level, the program

would randomly choose a learning task in this complexity level. All participants were unfamiliar with the task attributes of the ten complexity levels. Since the pre-training consisted of 2 tasks each from complexity levels 1 and 2, the participants in this condition could choose a learning task of complexity levels 3 to 10. The participants had a maximum jump size of 8 because they could pick a learning task of the highest complexity level (10) right after finishing the pre-training.

(4) Yoked preference condition. The same principle of the yoked efficiency condition applies for the yoked preference condition. Participants in the yoked preference condition were presented with a fixed training sequence. The individual training sequences of the participants in the personalized preference condition were divided over the participants in the corresponding yoked condition. Each training sequence was allocated only once to one participant in the yoked preference condition.

Dependent Variables

Training phase. First, the results of five dependent variables of the four experimental conditions will be given in order to gain insight in the task selection process. These variables are: number of learning tasks that was completed before reaching the highest complexity level, total number of learning tasks, training time, mean complexity level that was reached during training, and size of the jumps between complexity levels. Then, the results for performance and mental effort during training are provided.

Transfer test phase. After completion of the training, the participants were required to solve a twofold transfer test. First, their speed-accuracy on conflict identification was tested using a reaction time (RT) test in which screen dumps of ATC-situations were presented for 10 seconds. For the RT test, results will be given on two dependent variables: mean RT on conflict identification and the number of correct conflict identifications.

Second, they were required to solve ten transfer tasks that were structurally different from the learning tasks in several ways. Frequently, new aircraft appeared in the interface. Also, some of the aircraft had a different speed than other aircraft while all aircraft had the same speed in the learning tasks. Furthermore, the number of aircraft that had to be directed to their appropriate landing spot was larger than in any of the learning tasks. The results for performance and mental effort on these transfer tasks will be provided.

The efficiency measures of the four experimental conditions will be given using a new version of the efficiency formula as recently proposed by Tuovinen and Paas (2004) and Salden et al. (*Chapter 3:* 2004), in which training efficiency is calculated on the basis of three dimensions. The current study also adopts this

3D efficiency formula, using two test performance measures (reaction time test and transfer tasks) and one test effort measure, to determine training efficiency of the task selection methods. The training efficiency was determined using the following formula [1] (Tuovinen & Paas, 2004):

$$E = \frac{RT + P - ME}{\sqrt{3}}$$

In this formula, E = mental efficiency, ME = mental effort during test, RT = reaction time performance, and P = test performance. Analyses of variance (ANOVAs) and planned comparisons are used to analyze the data of the training and transfer test phases.

Results

Training Phase

Training Effects

No effects for the independent variable Method of Task Selection were found on the variables number of completed tasks before reaching the highest complexity level, total number of learning tasks, and training time (all Fs < 1). Means and standard deviations are provided in Table 4.

With regard to the mean complexity level that was reached during the training, a significant effect was found for Method of Task Selection, $F(3, 56) = 3.04$, $MSE = 2.07$, $p < .05$, $\eta^2 = .14$. With regard to the first hypothesis, no difference was found between the personalized conditions and the yoked conditions ($t < 1$). With regard to the second hypothesis, planned comparisons showed that the personalized efficiency condition reached a higher mean complexity level ($t(56) = 2.19$, $p < .05$) than the personalized preference condition.

The absolute jump size in complexity level depended on the learners' mental effort, performance, or mental efficiency on the previous task. Jumps in complexity level could be both negative and positive, leading to easier or more difficult tasks, respectively. In the analysis, the focus was on differences in absolute jump size. Using ANOVA, a main effect of Method of Task Selection was found, $F(3, 56) = 5.27$, $MSE = 0.03$, $p < .01$, $\eta^2 = .22$. With regard to the first hypothesis, no difference was found between the personalized conditions and

[1] If P (x_1, y_1, z_1) is a point in a 3-dimensional Cartesian space then the shortest distance, d, between it and the plane $x + y - z = 0$ is given by $d = (z_1 - x_1 - y_1) / \sqrt{3}$
The exact computations and argumentation for the 3-factor efficiency are presented by Tuovinen and Paas (2004).

Table 4 *Overview of training results.*

	Method of Task Selection							
	Efficiency		Yoked efficiency		Preference		Yoked preference	
Dependent variables [1]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
N tasks before highest complexity	4.73	3.41	4.80	3.39	6.25	4.29	6.17	4.37
Total *N* of learning tasks	10.20	2.76	10.00	2.67	12.13	4.22	12.00	4.41
Training time	1787.9	471.0	1823.9	482.4	2020.9	689.2	1982.1	668.7
Mean complexity	8.30	1.00	8.24	1.06	7.15	1.75	7.15	1.75
Jump size	0.90	0.10	0.92	0.10	0.72	0.23	0.73	0.23
Mental effort	3.3	0.4	3.1	0.6	3.0	0.7	3.1	0.5
Performance	81.09	9.52	68.67	10.59	74.97	10.20	73.44	16.40

[1] The slight differences on the five training phase variables between the personalized and its corresponding yoked condition are caused by a program failure which sometimes presented a participant with only 1 task of complexity level 2 during the pre-training.

the yoked conditions ($t < 1$). However, planned comparisons did support the second hypothesis because the personalized efficiency condition made larger jumps in complexity levels ($t(56) = 2.69$, $p < .05$) than the personalized preference condition.

Performance and mental effort. No effects of Method of Task Selection were found on the training variable mental effort ($F < 1$). A strong trend was found for the training variable performance ($F(3, 56) = 2.74$, $MSE = 144.01$, $p = .05$, $\eta^2 = .13$). With regard to the first hypothesis, the mean performance score of the personalized efficiency and personalized preference conditions ($M = 78.03$, $SD = 9.86$) was higher ($t(56) = 2.25$, $p < .05$) than the mean performance score of their corresponding yoked conditions ($M = 71.05$, $SD = 13.50$). Subsequent planned comparisons showed that this effect is only caused by the difference between the personalized efficiency condition and its corresponding yoked condition ($t(56) = 2.84$, $p < .01$). Furthermore, planned comparisons supported the second hypothesis, indicating that the personalized efficiency condition attained a higher performance score ($t(56) = 2.44$, $p < .05$) than the personalized preference condition. Means and standard deviations are provided in Table 4.

Transfer Test Phase

Reaction time test

No effect for the independent variable Method of Task Selection was found on the variable mean RT on conflict identification ($F < 1$). Means and standard deviations are provided in Table 5.

With regard to the number of correct conflict identifications, an ANOVA revealed a significant effect for the factor Method of Task Selection, $F(3, 56) = 8.18$, $MSE = 28.18$, $p < .0001$, $\eta^2 = .31$. Planned comparisons confirmed the first hypothesis, indicating that the mean number of correct conflict identifications ($M = 29.60$, $SD = 5.58$) of the personalized efficiency condition and personalized preference condition was higher ($t(56) = 2.04$, $p < .05$) than the mean number of conflict identifications ($M = 26.80$, $SD = 5.00$) of the yoked efficiency condition and the yoked preference condition. Furthermore, planned comparisons revealed the opposite effect of what was expected in the second hypothesis, namely that the personalized preference condition attained more conflict identifications ($t(56) = -3.58$, $p < .01$) than the personalized efficiency condition.

Transfer tasks

When analyzing the data on the transfer tasks, no differences were found between the different methods of task selection. There were no significant differences in mental effort or performance between the four experimental groups on the transfer tasks (all Fs < 1).

Table 5 *Overview of transfer test results.*

	Method of Task Selection							
	Efficiency		Yoked efficiency		Preference		Yoked preference	
Dependent variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RT Test								
Mean RT on conflict identification	6.64	1.15	6.08	1.51	6.06	1.48	6.62	1.00
N correct conflict identifications	26.13	5.84	24.13	4.70	33.07	5.32	29.47	5.30
Transfer Tasks								
Mental effort	3.9	0.5	3.6	0.6	3.7	0.8	3.6	0.5
Performance	64.79	6.30	62.46	11.08	67.13	8.49	65.79	10.28
Training efficiency	-0.43	0.62	-0.42	0.89	0.60	1.13	0.25	1.04

Training efficiency

Using an ANOVA, a significant effect was found for Method of Task Selection, $F(3, 56) = 4.45$, $MSE = 0.89$, $p < .01$, $\eta^2 = .19$. With regard to the first hypothesis, no difference was found between the personalized conditions and the yoked conditions ($t < 1$). In contrast with the second hypothesis, planned comparisons showed that the personalized efficiency condition was less efficient ($t(56) = -3.00$, $p < .01$) than the personalized preference condition. The means and standard deviations are provided in Table 5.

Discussion

The main hypothesis of this study that personalized dynamic task selection leads to more efficient training and better transfer performance than non-dynamic task selection was partially confirmed. While a strong trend showed that the mean training performance score of both personalized conditions was higher than the mean performance score of their corresponding yoked conditions, further analyses revealed that this was only caused by the difference between the personalized efficiency condition and its corresponding yoked condition. With regard to test performance, the mean number of correct RT conflict identifications of both personalized conditions was higher than the mean number of conflict identifications of their yoked conditions.

No support was found for the second hypothesis that task selection based on personalized efficiency would lead to more efficient training and higher transfer than selection based on personalized preference. Even though students in this condition reached a higher mean complexity level during training, made larger jumps during training, and a strong trend indicated a higher training performance, no beneficial effects were found during the test phase. Moreover, the RT test performance score of the personalized efficiency condition was lower than the performance score of the personalized preference condition. Also, the 3D efficiency formula showed that in terms of both test performances and test effort, the personalized efficiency condition was less efficient than the personalized preference condition. Overall, it can be concluded that the personalized efficiency condition did lead to superior effects during the training phase yet not during the test phase.

The small amount of differences between the personalized conditions and their associated yoked conditions might be attributed to the pre-training. While its purpose was to familiarize the participants with the program, the pre-training might have done more than that. It is plausible that due to this pre-training, the students were also enabled to enter the actual training with a similar amount of prior knowledge and acquired skills. It is likely that the pre-training has dimmed the possible range of effects that might have been found if the participants had entered the actual training directly after the introduction.

Another explanation for the small differences between the personalized conditions might be found in the different effects of these two conditions. While the personalized efficiency condition led to more training benefits than the other three conditions, the personalized preference condition led to higher transfer performance and training efficiency. Since both personalized methods did not lead to similar results, the claim that both dynamic conditions would lead to more efficient training and better transfer performance than the non-dynamic conditions, was not fully supported.

Furthermore, it might also be that the personalized conditions lead to individual training sequences that were useful for most participants in the yoked conditions. Despite the fact that the sequences were not personalized for these participants, the progress the participants made to which they were linked was good enough to follow through. Even though a participant in a yoked condition might have chosen a learning task of a different complexity level at a certain point in training, usually, s/he could cope with the predetermined task order and rise in complexity levels. It is likely that our participants were homogeneous in their prior knowledge, which means that they might have chosen more or less the same training sequence. With more variation in learners' expertise levels, larger differences in the training sequences would have been expected.

When taking a closer look at the effects of the personalized efficiency condition, it seems that in this condition, the participants achieved the highest mean complexity level and made the largest jumps between complexity levels. Furthermore, a strong trend can be observed which shows that the efficiency condition attained the highest performance score on the training while investing an equal amount of mental effort as the other conditions. However, despite these training benefits, no effects occurred on the two transfer tests and the efficiency formula.

A possible explanation for the small beneficial effects of the personalized efficiency condition can be found in the fact that task selection was only based on complexity level, but not on the type of conflict the participants had trouble with. In one complexity level, different types of conflict can occur and it could be that a participant who made a single mistake concerning one type of conflict would be presented with a more complex learning task. However, s/he would not have the chance to practice coping with this type of conflict again on the same complexity level but only in the context of more complex learning tasks. If task selection is specified on the level of types of conflicts then the participants will receive more practice on the type of conflict resolution they had trouble with. Such further refinement of the task selection process in future research might lead to more beneficial effects for the personalized efficiency method.

Also, it would be interesting to administer a delayed transfer test some time (e.g., two weeks) after the training was given. If either personalized

method would lead to deeper processing and better storage of the learning material, delayed testing might be able to reveal this. If the personalized preference method would lead to more elaborated schemata, then the participants trained with this method should exhibit higher performance than the participants trained with the efficiency method. Besides administering transfer tasks to see whether the delayed testing reveals any effects, it would also be interesting to examine whether the initial effect of the RT test can still be detected.

In contrast to various studies on learner control (e.g., Bell & Kozlowski, 2002; Niemic, Sikorski, & Walberg, 1996; Steinberg, 1977, 1989; Williams, 1993), the present study shows that students are capable of using the given control when the training situation is well constructed and avoids overloading their cognitive system. The finding that novices are able to deal with a certain degree of freedom, makes it particularly interesting to investigate to what extent more expert participants are able to use learner control. Using students that already have acquired a certain level of expertise, one could experiment by increasing the degrees of freedom. With a higher expertise level, it can be expected that the student would benefit most from a large amount of given freedom.

The use of the extended 3D efficiency formula of Tuovinen and Paas (2004) proved to be successful. It can be used flexibly as it enables one to take more important variables into account that might differ from experiment to experiment. In the current study, transfer performance scores were collected and used in the 3D formula. This led to a realistic efficiency score that represents a complete view on the transfer phase which would not have been possible when using the original 2D version of the formula of Paas and van Merriënboer (1993). Even though no effects were found on the transfer tasks in terms of performance or invested mental effort, the effect on the RT test appeared strong enough to create significant efficiency effects.

In conclusion, the results regarding the first hypothesis of this study partially supported the idea that adapting training to the individual needs of the student can make training more efficient. No evidence was found to support the second hypothesis, which stated that task selection based on personalized efficiency can lead to more efficient training and higher transfer than selection based on personalized preference. While the personalized efficiency condition showed several training benefits, it did not prove to lead to higher transfer performance nor was this condition more efficient. In contrast, the personalized preference condition showed only minor training benefits, yet it did lead to higher transfer performance and was shown to be an efficient training method.

The current combined research on the efficiency method so far, has shown that future research on the efficiency method is needed to fully grasp its benefits and shortcomings. While in previous studies this method was at least as good as other dynamic methods, it has been shown that it could not compete

with the personalized preference condition in terms of transfer performance and training efficiency. In contrast to various studies on learner control, the current study has shown that students are capable of handling given control of training, as long as their cognitive systems are not overloaded. This has implications for future research in which students of varying levels of expertise can be given learner control to varying degrees of freedom.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science, 205*, 1289-1292.

Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology, 55*, 267-306.

Borsook, T. K., & Higginbotham-Wheat, N. (1991). Interactivity: what is it and what can it do for computer-based instruction? *Educational Technology, 4*, 11-17.

Camp, G., Paas, F., Rikers, R. M. J. P., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Crooks, S. M., & Klein, J. D. (1996). Effects of cooperative learning and learner-control modes in computer-based instruction. *Journal of Research on Computing in Education, 29*, 109-124.

EATMP Human Resources Team (2001). *Guidelines for selection procedures and tests for ab initio trainee controllers (Revised)*, HRS/MSP-002-GUI-01. Brussels, Belgium: EUROCONTROL.

Galster, S. M., Duley, J. A., Masalonis, A. J., & Parasuraman, R. (2001). Air Traffic Controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology, 11*, 71-93.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5-23.

Kashihara, A., Hirashima, T., & Toyoda, J. (1995). A cognitive load application in tutoring. *Journal of User Modeling and User-Adapted Instruction, 4*, 279-303.

Niemic, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*, 157-174.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Tuovinen, J. E., Tabbers, H. K., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Reeves, T. C. (1993). Pseudoscience in computer-based instruction: The case of learner control research. *Journal of Computer-Based Instruction, 20*, 39-46.

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science, 32*, 153-172.

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

Steinberg, E. R. (1977). Review of student control in computer-assisted instruction. *Journal of Computer-Based Instruction, 3*, 84-90.

Steinberg, E. R. (1989). Cognition and learner control: A literature review, 1977-1988. *Journal of Computer-Based Instruction, 16*, 117-121.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*, 133-152.

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance, and training efficiency. *Learning and Instruction, 12*, 11-37.

Williams, M. D. (1993). A comprehensive review of learner-control: The role of learner characteristics. In M. R. Simonson & A. Dristen (Eds.), *Proceedings of the Annual Conference of the Association for Educational Communications and Technology* (pp. 1083-1114). New Orleans, LA: Association for Educational Communications and Technology.

Chapter 5 – Dynamic Task Selection in Flight Management System Training [1]

Abstract

The effects of three task selection methods on test performance and training efficiency were investigated in a computer-based Flight Management System (FMS) training. A fixed condition was compared to a learner control condition, and a condition using the participants' self-rated performance and mental effort for task selection. Although the experimental conditions revealed more positive training effects, no differences were found for training efficiency and test performance. A follow-up study did not confirm the alternative hypothesis that these results were caused by the higher amount of tasks in the fixed condition. Additional analyses suggested that the quality of self-rating needs to be considered in future research.

Introduction

In the aviation domain continuous efforts are directed at increasing cockpit automation. An automated cockpit has a significant effect on the pilot's tasks and demands additional competencies from the cockpit crew. The Flight Management System (FMS) is one of the core systems in an automated cockpit, which can control an entire flight from take off to landing. Considering the importance of the FMS it is remarkable that automation was introduced relatively late in training programs and that until recently, realistic computer-based FMS simulations were rarely available. Computer-based training may prepare and enhance the pilot's automation related skills and make time spent on expensive part-task trainers and full flight simulators more effective. Whereas, it may allow experienced pilots to practice on new FMS systems in a free-play fashion, novice pilots can be given more support, for instance, by adapting the complexity of learning tasks to their experience level. It can be expected that such personalized training can make FMS training more efficient.

This study examines several training methods that were designed according to the 4C/ID-model (van Merriënboer, 1997; van Merriënboer, Clark, & de Croock, 2002). This model offers a training design that presents students with a predetermined order and complexity of learning tasks in such a way that their cognitive capacity is optimally used. Further efficiency may be reached by providing a personalized and adaptive training trajectory, in which learning tasks are selected during training based upon the performance and needs of the individual learner. Especially for training complex cognitive skills, the use of such a 'dynamic' ability to optimize task order and complexity for the individual trainee is believed to be strongly related to increased training

[1] *This chapter is based on: Salden, R. J. C. M., Paas, F., van der Pal, J., & van Merriënboer, J. J. G. (2004). Dynamic task selection in a Flight Management System training. Manuscript submitted for publication.*

efficiency (*Chapter 3:* Salden, Paas, Broers, & van Merriënboer, 2004; *Chapter 2:* Salden, Paas, & van Merriënboer, in press).

Research in the context of cognitive load theory (for an overview see Paas, Renkl, & Sweller, 2003, 2004) has shown that cognitive load is a crucial factor in the training of complex cognitive skills. The combination of cognitive load and performance measures is considered a superior estimate of a learner's cognitive demands that can be used in the dynamic selection of learning tasks. For example, when two trainees achieve the same performance scores, measures of cognitive load might be able to reveal differences in training efficiency otherwise unnoticed. While one of them might have experienced a very high cognitive load and needed to work laboriously through a very effortful process, the other person might have experienced a low cognitive load and reached the same performance level with a minimum of effort. In a personalized task selection method this information could be used to present a less difficult next task to the first person than to the second person. Because the combination of performance and cognitive load measures provides a clear picture of the state of the student's cognitive system at a certain moment in training, dynamically selected learning tasks fit well to the cognitive schemata a student has acquired. The individual capacity of a student is taken into account, leading to high training efficiency. To obtain a good indication of the cognitive load that is imposed on a person's cognitive system, mental effort measurements may be used (Paas & van Merriënboer, 1993).

Paas and van Merriënboer (1993; see also Paas, Tuovinen, Tabbers, & van Gerven, 2003) have developed a calculational approach for combining measures of mental effort and task performance that allows one to obtain information on the relative efficiency of instructional conditions. Based on Ahern and Beatty's (1979) efficiency view on learning, it is proposed that learners' behavior in a certain training condition is more efficient if (1) their performance is higher than might be expected on the basis of their invested mental effort, and/or (2) their invested mental effort is lower than might be expected on the basis of their performance. Thus, training conditions in which high performance is attained with a low mental effort investment are considered as 'high efficient'. 'Low efficient' conditions are characterized by a combination of low performance and high mental effort.

A first indication that the use of a combined performance and mental effort score can make personalized training more efficient was found in two studies of Camp, Paas, Rikers, and van Merriënboer (2001), and Salden et al. (*Chapter 3:* 2004). Both studies compared four methods of task selection in the Air Traffic Control (ATC) domain. In the first method, tasks were presented in a fixed, predetermined simple-to-complex sequence designed according to the 4C/ID-model. In the other three methods, the tasks were presented dynamically, based on either performance, mental effort, or the combination of

both (i.e., mental efficiency). Results showed that dynamic task selection leads to more efficient training than non-dynamic task selection. However, dynamic task selection based on mental efficiency did not lead to more efficient training and better test performance than dynamic task selection based on performance or mental effort alone.

Besides automated task selection, learner control may offer another form of dynamic task selection because it gives the students control over what learning tasks they want to practice next. The theoretical claim for learner control is that trainees are capable of selecting appropriate tasks to practice and can avoid a possible overload of their cognitive system, thereby increasing the effectiveness and efficiency of learning (e.g., Borsook & Higginbotham-Wheat, 1991; Niemic, Sikorski, & Walberg, 1996; Steinberg, 1977, 1989; Williams, 1993). However, according to Bell and Kozlowski (2002) positive effects of learner control on learning can only be expected if instructional materials are designed in such a way that they provide learners with a level of control they are able to handle. Support for this claim was found in recent studies (van Merriënboer, Schuurman, de Croock, & Paas, 2002; *Chapter 4:* Salden, Paas, & van Merriënboer, 2004), which showed that learners who are given an appropriate level of control over task selection are well able to select their own learning tasks.

In the current study, participants with some pilot background but no FMS knowledge are divided into three conditions of FMS training. The fixed condition is a control condition in which learning tasks are presented in a predetermined order based on increasing complexity of learning tasks. This condition will be compared with two personalized experimental conditions, in which learners either have to select the learning task themselves, i.e. the learner control condition, or in which the learning tasks are selected by the training program using a combination of the learners' self-rated performance and mental effort, i.e. the mental efficiency condition.

In agreement with Camp et al. (2001) and Salden et al. (*Chapters 3 and 4:* 2004), it is hypothesized that personalized task selection leads to more efficient training and better test performance than non-dynamic task selection. The differences between both dynamic conditions, which personalize the task selection either by using the mental efficiency or by allowing the learners to determine their own training sequence, are explored.

Experiment 1

Method

Participants

Thirty-one students of a higher education school for aviation (3 women and 28 men, *M* = 20.1 years, *SD* = 2.69), who were novices with regard to the FMS, participated in this study. The students were randomly assigned to three experimental conditions: A fixed condition (*n* = 10), a learner control condition (*n* = 10), and a mental efficiency condition (*n* = 11). All participants were in good health and had normal or corrected-to-normal vision. They volunteered to participate in this study.

Materials

FMS simulation. The training software was based on a realistic computer simulation of a Boeing 747 FMS developed by the National Aerospace Laboratory NLR. The training software ran on an IBM-compatible PC (Pentium III, 533 MHz) using an IBM 17-inch SVGA monitor. Figure 1 depicts the interface of the FMS program.

Figure 1. Interface of the FMS training program.

Learning tasks. In the training, the participants were confronted with learning tasks, which presented flight information of a certain route from airport A to airport B that learners had to program into the FMS simulation. A simulated flight had to be executed after entering all information. At certain points during the task, changes in the flight route were required and made it necessary for the trainees to adjust the original flight route. Possible changes consisted of an alteration in arrival data (e.g., a new Standard Terminal Arrival Route), a new runway, or a diversion to another airport.

Prior to training, thirty-two learning tasks were categorized into eight levels (four tasks per level) that specified the complexity of the tasks based on three complexity factors: The amount of data to be programmed into the FMS program, the number of changes in flight route, and the amount of time pressure. Values on these factors were added to determine the overall complexity of a learning task.

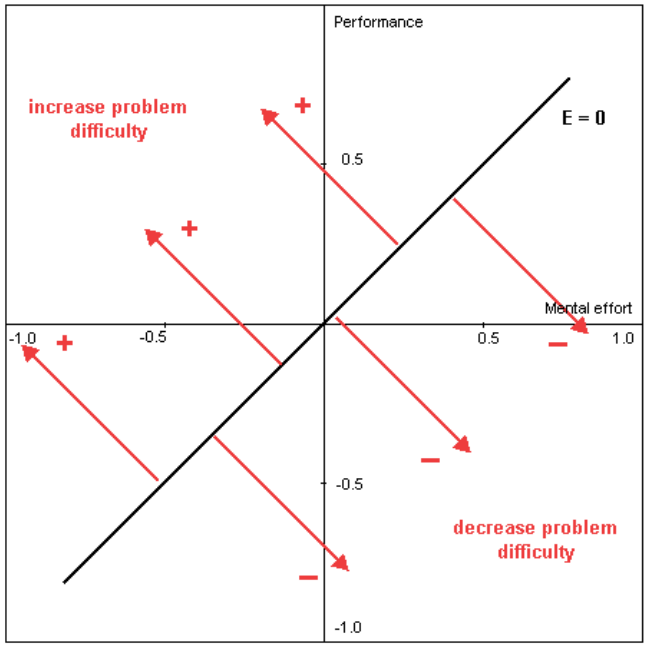

Figure 2. Representation of the effect of mental efficiency on the selection of the complexity of the next learning task.

Task selection. The selection of tasks differed between the three experimental conditions. In the fixed condition, participants received a total of 16 learning tasks with two randomly chosen tasks of each of the eight complexity levels (1, 1, 2, 2, 3, 3, etc.). These 16 learning tasks were presented in a predetermined order from low to high complexity. In the learner control condition, participants received an overview of all learning tasks with an

indication of their complexity and could choose which task to practice next. Thus, these learners had maximum freedom to determine their own training sequence.

In the mental efficiency condition, task selection was based on participants' self-ratings of performance and mental effort on a 5-point rating scale. These subjective performance and mental effort scores were used to determine mental efficiency (Paas & van Merriënboer, 1993; Paas et al., 2003). If the subjective performance score (see Y axis of Figure 2) was higher than the mental effort score (X axis) this was interpreted as a high mental efficiency; the learner performed relatively well and invested less mental effort than could be expected on the basis of his or her subjective performance. If the subjective performance score was lower than the mental effort score this was interpreted as a low mental efficiency; the learner performed relatively low and invested more mental effort than could be expected on the basis of his or her subjective performance. As indicated in Table 1, the complexity of the next learning task was determined on the basis of this argumentation. The subjective performance and mental effort scores in this Table correspond with the self-ratings of the individual student. For instance, if a participant had a mental effort score of 2 and a subjective performance score of 5, task complexity was increased with three levels (+3). Note that participants in the mental efficiency condition had a maximum possible step size of 4 complexity levels, while participants in the learner control condition had a maximum possible step size of 7 complexity levels (i.e., directly from the lowest to the highest complexity level or vice versa).

Table 1 *Selection table indicating jump size in complexity between learning tasks.*

Mental effort	Performance				
	1	2	3	4	5
1	0	+1	+2	+3	+4
2	-1	0	+1	+2	+3
3	-2	-1	0	+1	+2
4	-3	-2	-1	0	+1
5	-4	-3	-2	-1	0

Three possible outcomes could lead to the completion of the training. First, participants finished practice after working on 16 learning tasks. This number of tasks ensures sufficient variation over the eight complexity levels. Second, participants finished practice after their self-ratings equaled the preset performance (≥ 3) and mental effort (≤ 3) criteria for two successive learning tasks at the highest complexity level. Third, participants finished practice after they performed all four available learning tasks at the highest complexity level. Note that in the two latter cases a participant could complete the training after working on less than 16 learning tasks.

Test tasks. After the training, the participants had to perform a test that consisted of five test tasks, which were different from the learning tasks in two ways: (1) the amount of data that had to be programmed into the FMS simulation was increased, and (2) the number of changes in the flight route was higher.

Objective scoring of performance. To be able to compare the training and test performance between the experimental conditions, the objective performance of all participants was scored after completion of the experiment. For each flight, scores on a 5-point scale were given on four performance variables: (a) the given number of commands; (b) the number of correct commands; (c) the time on task, and (d) the time needed to process a change in flight route data. For all variables, a score of 1 indicated a very low performance and a score of 5 indicated a very high performance. The mean score of these four variables was used to compare performance between the experimental conditions.

Procedure

All participants were given a paper-based 10-pages introduction to the training and the use of the FMS simulation, which presented all information required for the training, including examples of how to enter commands into the FMS. Participants were free to consult this introduction during the entire training session. After they had read the introduction, the training started and a learning task at the first complexity level was presented. The subsequent tasks depended on the experimental condition. In the fixed condition, participants received another task at complexity level 1, then two tasks at level 2, and so forth; in the learner control condition, the participants could choose whatever task they preferred next; and in the mental efficiency condition, the next task was based on their subjective performance and mental effort self-ratings (see Table 1 for the applied step size). All participants could continue with the next learning task as soon as they had completed the previous task, meaning that differences in training time could occur in all conditions. The participants performed the five test tasks immediately after they completed the learning tasks. The whole experiment took about three hours.

Results and Discussion

First, the results on the training phase are reported. The mean number of learning tasks, step sizes, and total training time are given for each condition to provide insight in the task selection process. Furthermore, the results for training performance and mental effort are given. Second, the results for performance, mental effort, and training efficiency are provided for the test phase. One-sample and independent t-tests, ANOVAs, ANCOVAs and planned comparisons were used to analyze the data. Means and standard deviations are presented in Table 2.

Training Phase

Training effects. Because the number of learning tasks was preset at 16 in the fixed condition, one-sample t-tests were used to compare this number of tasks to those of the learner control and mental efficiency conditions. Both the learner control condition (M = 6.50, SD = 1.35; $t(19)$ = -4.3, $p < .001$) and the mental efficiency condition (M = 7.27, SD = 1.19; $t(20)$ = -4.6, $p < .001$) needed substantially less than the 16 learning tasks in the fixed condition to complete the training. The comparison between the learner control and mental efficiency condition showed no difference in number of learning tasks (t = 1.4).

Table 2 *Overview of the results of Experiment 1.*

	Method of Task Selection					
	Fixed		Learner control		Mental efficiency	
Dependent variables	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training phase						
N of learning tasks	16	-	6.50	1.35	7.27	1.19
Step size	7/15	-	.95	.07	.93	.09
Training time	149.60	22.77	78.69	1.64	117.35	25.61
Mental effort	2.24	.52	2.13	.56	2.38	.41
Performance	3.27	.28	2.64	.38	2.70	.18
Test phase [a]						
Mental effort	2.41	.93	1.61	.53	2.22	.49
Performance	2.89	.38	3.16	.22	3.21	.20
Training efficiency	-1.09	1.17	.62	.67	.43	.61

The absolute step size in complexity level was also preset for the fixed condition, at one level in complexity per two tasks. In total the participants made 15 steps between 16 tasks, with 7 steps to a higher complexity level (M = .47 (7/15)). For the learner control and mental efficiency condition, steps between complexity levels could be negative or positive, corresponding to easier or more difficult tasks, respectively. One-sample t-tests on the absolute step size showed that participants in both the learner control condition (M = .95, SD = .07; $t(19)$ = 4.1, $p < .01$) and the mental efficiency condition (M = .93, SD = .09; $t(20)$ = 4.3, $p < .001$) made larger absolute steps between complexity levels than participants in the fixed condition. The step sizes between the learner control and mental efficiency condition did not differ ($t < 1$).

A significant effect was found for total training time, $F(2, 28) = 28.37$, MSE = 444.40, $p < .001$, $\eta^2 = .67$. Planned comparisons showed that the participants in the fixed condition (M = 149.60, SD = 22.77) needed more time to complete the

[a] Estimated marginal means are presented with number of learning tasks and total training time as covariates.

training ($t(28) = 6.37$, $p < .001$) than the participants in the learner control and mental efficiency conditions ($M = 98.02$, $SD = 18.63$). Furthermore, the participants in the learner control condition needed less training time than those in the mental efficiency condition ($t(28) = -4.20$, $p < .001$).

Performance and mental effort. A significant effect was found for training performance ($F(2, 28) = 15.00$, $MSE = .08$, $p < .001$, $\eta^2 = .52$). The performance score of the fixed condition ($M = 3.27$, $SD = .28$) was higher ($t(28) = 5.47$, $p < .001$) than the mean performance score of the learner control and mental efficiency conditions ($M = 2.67$, $SD = .28$). The comparison between the learner control and mental efficiency condition showed no difference in performance ($t < 1$). No significant effects were found on the mental effort during training ($F < 1$).

Test Phase

During the training, participants in the fixed condition worked on many more tasks, made smaller steps between complexity levels, and needed more training time than participants in the learner control and mental efficiency conditions. Especially the number of tasks and the total training time could easily explain possible differences between conditions on the test tasks. Therefore, number of learning tasks and total training time are included as covariates in the subsequent analyses.

Performance and mental effort. Using an ANCOVA with number of learning tasks and total training time as covariates, no effects were found on performance and mental effort (Fs < 1).

Training efficiency. The training efficiency was determined using the following formula (Paas et al., 2003; Paas & van Merriënboer, 1993):

$$E = \frac{P - ME}{\sqrt{2}}$$

In this formula, E = mental efficiency, P = test performance, and ME = mental effort during training.

Using an ANCOVA with number of learning tasks and total training time as covariates, no effect was found for Method of Task Selection, ($F < 1$). The estimated marginal means and standard deviations are provided in Table 2.

In conclusion, participants in the learner control and mental efficiency conditions worked on less learning tasks, made larger steps between complexity levels and needed less time to complete the training than the participants in the fixed condition. Performance during training was higher for the fixed condition, which can easily be explained by the prolonged training time. No differences were found on test performance, mental effort during the test phase, and training efficiency.

To control for the high number of learning tasks in the fixed condition (16) a second experiment was conducted, comparing a mental efficiency

condition to a fixed condition with only 8 learning tasks. It was expected that this would limit the difference between conditions during the training phase, and possibly show the expected positive effect of dynamic task selection on test performance.

Experiment 2

Method

Participants

Twenty students of the same higher education school for aviation as in the first experiment (6 women and 14 men, M = 23.8 years, SD = 4.12), who were novices with regard to the FMS, participated in this study. The students were randomly assigned to the two experimental conditions: A fixed condition (n = 10) and a mental efficiency condition (n = 10). All participants were in good health and had normal or corrected-to-normal vision. They received € 15 (approximately $ 20) for their participation.

Materials and Procedure

The materials were the same as in Experiment 1. The only difference is that the number of learning tasks in the fixed condition was downsized from 16 to 8, resulting in only one task per complexity level. The procedure was identical to the procedure in Experiment 1 and participants received the same test tasks.

Results and Discussion

First, the results on the training phase are reported. The mean number of learning tasks, step sizes, and total training time are given for each condition to provide insight in the task selection process. Furthermore, the results for training performance and mental effort are given. Second, the results for performance, mental effort, and training efficiency are provided for the test phase. One sample t-tests, ANOVAs, and ANCOVAs were used to analyze the data. Means and standard deviations are presented in Table 3.

Training Phase

Training effects. Because the number of learning tasks was preset at 8 in the fixed condition, one-sample t-tests were used to compare the number of tasks with the mental efficiency condition. The t-test showed that the mental efficiency condition (M = 6.28, SD = 1.48; $t(19) = -2.9$, $p < .01$) needed less than the 8 learning tasks in the fixed condition to complete the training.

The absolute step size in complexity level was also preset for the fixed condition, at one step in complexity per task. In total the participants made 7

steps between 8 tasks (*M* = .88 (7/8)). For the mental efficiency condition, steps between complexity levels could be negative or positive, corresponding to easier or more difficult tasks, respectively. A one-sample t-test showed that the participants in the mental efficiency condition (*M* = .96, *SD* = .44; *t*(19) = 3.6, *p* < .01) made larger steps between complexity levels than those in the fixed condition. With regard to total training time, no effect was found for Method of Task Selection, (*F* = 1.29).

Table 3 *Overview of results Experiment 2.*

	Method of Task Selection			
	Fixed		Mental efficiency	
Dependent variables	*M*	*SD*	*M*	*SD*
Training phase				
Total N of learning tasks	8	-	6.20	1.48
Jump size	7/8	-	.96	.44
Training time	79.97	18.32	70.71	18.17
Mental effort	2.40	.46	2.27	.46
Performance	3.26	.24	3.40	.34
Test phase [a]				
Mental effort	1.76	.18	2.08	.18
Performance	3.45	.15	3.14	.15
Training efficiency	.46	.30	-.46	.30

Performance and mental effort. No effects of Method of Task Selection were found on the training variables mental effort (*F* < 1) and performance (*F* = 1.19).

Test Phase

An ANCOVA with number of learning tasks as a covariate showed no effects on mental effort (*F* = 1.05) and performance (*F* < 1). An ANCOVA with number of learning tasks as a covariate showed no effect (*F* = 2.67) on training efficiency. The estimated marginal means and standard deviations are provided in Table 3.

Additional Analyses for Experiment 1 and 2 Combined

Experimenter's observations of the participants in the mental efficiency conditions of both experiments suggested that the absence of clear beneficial effects for this condition might have been caused by the poor quality of self-ratings of performance (Bjork, 1999; Tousignant & DesMarchais, 2002). In particular, it seemed that some of the participants overrated their performance as compared to their objective performance scores.

[a] Estimated marginal means are presented with number of learning tasks as a covariate.

To test this alternative hypothesis, a K-means cluster analysis ($F(2, 18) = 71.6$, $MSE = .03$, $p < .001$) was performed on the differences between objective and subjective performance scores. Three groups of self-raters were identified: Good self-raters ($n = 6$), average self-raters ($n = 9$), and bad self-raters ($n = 6$). The extreme groups (i.e., good and bad self-raters) were compared to the combined fixed conditions of both studies on the test variables mental effort and performance.

Kruskal-Wallis tests revealed that the participants in the fixed condition ($M = 3.41$, $SD = .46$) attained a higher test performance than the bad self-raters ($\chi^2 = 7.21$, $p < .01$; $M = 2.89$, $SD = .24$). However, no difference was found between the fixed condition and the good self-raters ($\chi^2 < 1$; $M = 3.27$, $SD = .22$). In addition, the good self-raters attained a higher test performance ($\chi^2 = 5.04$, $p < .05$) than the bad self-raters. No effects were found on the test variable mental effort ($F = 1.8$). The means of the fixed group and the good self-raters and bad self-raters in the mental efficiency group are depicted in Figure 3.

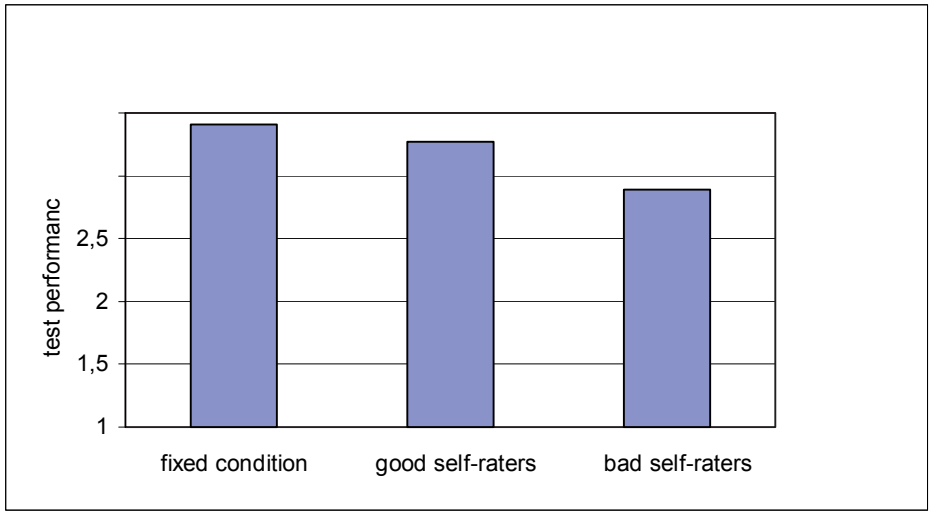

Figure 3. Histogram of test performance of fixed condition, good and bad self-raters in the mental efficiency condition.

General Discussion

The main hypothesis of the first experiment that dynamic task selection leads to more efficient training and better test performance than non-dynamic task selection was not confirmed. Although the participants in the learner control and mental efficiency conditions progressed through training more effectively than the participants in the fixed condition, training performance was highest in the fixed condition. This effect was explained by the prolonged training time. Using the number of learning tasks and total training time as covariates, no

effects were found for test performance, mental effort on the test, and training efficiency. Participants in the learner control condition resembled the participants in the mental efficiency condition to a high extent. The only difference found between these two conditions was that the participants in the learner control condition required less time to complete the training than the mental efficiency condition.

The reduction of the number of learning tasks in the second experiment was expected to show the positive effect of dynamic task selection on test performance. The participants in the mental efficiency condition still needed fewer tasks to complete training and made larger jumps than the participants in the fixed condition. No effects in support of the mental efficiency condition were found on mental effort, test performance and training efficiency.

From observing the participants in the mental efficiency conditions in both experiments, it was hypothesized that the absence of clear beneficial effects for this condition might have been caused by the low quality of the self-ratings. In other words, some of the participants would overrate their own performance compared to their actual objective performance. When comparing the fixed conditions to two subgroups of self-raters, effects were found on test performance. It was shown that the participants in the fixed condition attained a higher performance than the bad self-raters but no difference was found between the participants in the fixed condition and the good self-raters. Furthermore, the good self-raters attained a higher test performance than the bad self-raters. Based on these results, it seems plausible that the low quality of the self-ratings has confounded the results in both experiments. Since most participants were not very skilled at rating their own performance, the personalized training sequence they were presented with was not optimal. The overrating of their performance led them to receive more difficult tasks than they should have been given.

In addition to the confounding effects of the self-ratings, there are some possible explanations that can be given for the lack of beneficial effects of personalized task selection. For instance, it can be observed that the performance scores for all conditions were relatively high. Although additional analyses revealed no ceiling or floor effects, an alternative explanation is that the range of complexity used in this study was too limited. This suggestion is further strengthened by the relatively low levels of invested mental effort in all conditions.

The basic operations on the FMS are very recurrent since one always has to give certain commands in order to be able to execute a flight. The complexity in use can be increased when one has to deal with the FMS in a more authentic cockpit situation. The interaction of the FMS with many displays and control panels is important to achieve situational awareness for the trainee pilot. While the aim of our training was to familiarize the participants with the FMS, the

scope of the training may have been too limited. In accordance with the previous line of reasoning a larger range of complexity in the materials might have resulted in larger differences in performance and mental effort on both training and test phases.

Another explanation could be found in the use of the 4C/ID-model. Like in previous experiments (Camp et al., 2001; *Chapter 3:* Salden et al., 2004) this model was used to create the fixed conditions. The design guidelines in this model recommend a steady increase in complexity during training. While participants in the personalized conditions often complete training faster, only few beneficial effects were found on test performance and training efficiency. Whereas previous studies on Air Traffic Control (ATC) training were able to identify at least some beneficial effects, the combination of the 4C/ID-model and the recurrent nature of the FMS skills might deem personalized training methods unnecessary.

In conclusion, the results only moderately supported the idea that adapting training to the individual needs of the student makes training more efficient. While the mental efficiency condition proved to be an effective training method, it did not prove to be efficient as well. In contrast, the fixed condition proved not to be the most effective yet did prove to be a fairly efficient training method. The current combined research on the efficiency method so far, has shown that future research on the efficiency method is needed to fully grasp its benefits and shortcomings.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science, 205*, 1289-1292.

Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology, 55*, 267-306.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: The MIT Press.

Borsook, T. K., & Higginbotham-Wheat, N. (1991). Interactivity: what is it and what can it do for computer-based instruction? *Educational Technology, 4*, 11-17.

Camp, G., Paas, F., Rikers, R. M. J. P., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Niemic, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*, 157-174.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paas, F., Tuovinen, J. E., Tabbers, H. K., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science, 32*, 153-172.

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

Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (2004). *Personalized task selection in Air Traffic Control: Effects on training efficiency and transfer*. Manuscript submitted for publication.

Steinberg, E. R. (1977). Review of student control in computer-assisted instruction. *Journal of Computer-Based Instruction, 3*, 84-90.

Steinberg, E. R. (1989). Cognition and learner control: A literature review, 1977-1988. *Journal of Computer-Based Instruction, 16*, 117-121.

Tousignant, M., & DesMarchais, J. E. (2002). Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: A correlation study. *Advances in Health Sciences Education, 7*, 19-27.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology, Research and Development, 50*, 39-64.

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance, and training efficiency. *Learning and Instruction, 12*, 11-37.

Williams, M. D. (1993). A comprehensive review of learner-control: The role of learner characteristics. In M. R. Simonson & A. Dristen (Eds.), *Proceedings of the Annual Conference of the Association for Educational Communications and Technology* (pp. 1083-1114). New Orleans, LA: Association for Educational Communications and Technology.

Chapter 6 – General Discussion

The main aim of this dissertation was to investigate the use of personalized instruction in the training of complex cognitive skills. Based on a comparison of learning task selection methods in Chapter 2, it was shown that cognitive load (for an overview see Paas, Renkl, & Sweller, 2003, 2004) is a crucial factor in the training of complex cognitive skills. The combination of cognitive load and performance measures is considered a superior estimate of a learner's cognitive demands that can be used in the dynamic selection of learning tasks. While two students might achieve the same performance score they might have experienced different levels of cognitive load, which should be taken into account in the task selection process to ensure an optimal learning process for each individual student.

Consequently, the four studies in this dissertation aimed at investigating dynamic task selection methods based on several variables such as performance, cognitive load, and a combination of both (i.e., mental efficiency). It was hypothesized that dynamic whole-task methods could be enhanced by using the combination of performance and mental effort scores (i.e., mental efficiency) in the process of task selection. The four studies in Chapters 3 through 5 examined the mental efficiency method (Paas & van Merriënboer, 1993, 1994a, 1994b) in various instructional settings and in two training fields of the aviation domain: Air Traffic Control (ATC) and Flight Management Systems (FMS). While automated systems are used more and more frequently in the training of such complex cognitive skills, the selection of training tasks is still not automatically adapted to learner characteristics. The first study to use the mental efficiency method for dynamic task selection in the aviation domain was conducted by Camp, Paas, Rikers, van Merriënboer (2001). Although the results of this study showed an overall beneficial effect of dynamic task selection as compared to non-dynamic task selection, no beneficial effect for the efficiency method was found. The studies in this dissertation used the Camp et al. study as a starting point.

This final chapter briefly reviews the results of these studies, discusses the limitations of the experiments, and describes the implications of the results for future research.

Review of results

The comparison of learning task selection methods in Chapter 2 showed an evolution in training programs in three important ways: (a) from non-dynamic to dynamic, (b) from part-task based to whole-task based, and (c) from group-based to personalized. Especially, the use of dynamic and personalized selection of whole tasks is believed to be strongly related to increased training

efficiency. Furthermore, research in the context of cognitive load theory (for an overview see Paas, et al., 2003, 2004) was presented, indicating that cognitive load is a crucial factor to consider for the training of complex cognitive skills.

The first ATC study (Chapter 3) was a partial replication of Camp et al. (2001) and compared a non-dynamic condition, in which the learning tasks were presented to the participants in a fixed predetermined sequence, to three dynamic conditions, in which learning tasks were selected on the basis of, in order, performance, mental effort, and mental efficiency. Several changes were made in the materials. First, with regard to the selection algorithm, the maximum jump size between complexity levels was decreased from four in the Camp et al. study to two in the present study, forcing a smoother increase or decrease in task complexity. Furthermore, while all participants in the Camp et al. study had to practice 20 learning tasks before being able to continue to the test tasks, our participants could continue to the test tasks as soon as they had achieved criterion scores on training tasks of the highest complexity level. Lastly, the number of test tasks was increased. Again, the results showed a beneficial effect for dynamic task selection, but no beneficial effect for the efficiency condition in comparison to the other two dynamic conditions (i.e., based on performance or mental effort).

In the second ATC study (Chapter 4), two personalized methods were contrasted to yoked control conditions. The personalized efficiency and learner preference (learner control) conditions showed superior results on a reaction time test, yet no difference was found on the transfer tasks. While the personalized efficiency condition showed more training benefits in comparison to the personalized learner preference condition, the latter condition proved to be more efficient than the personalized efficiency condition.

The last two studies (Chapter 5) focused on another task within the aviation domain, namely, operating a Flight Management System. The third study compared a non-personalized training sequence to a learner control condition in which learners were free to select a new learning task, and a mental efficiency condition in which participants had to self-rate their mental effort and performance on the basis of which a new task was selected by the system. Results showed that while both personalized conditions attained better training effects, no differences could be found on the performance of the test tasks. The fourth study investigated whether the higher amount of training time and the larger number of training tasks in the non-personalized condition confounded the results. However, despite a reduction of training time by reducing the amount of training tasks in the non-personalized condition, the efficiency condition did not exhibit superior performance on the test. Additional cross-study analyses revealed an important difference between good and bad self-raters, which might have confounded possible beneficial effects of the efficiency method.

The results of the four studies lead to the following conclusions. First of all, personalized instruction can have beneficial effects for the training of complex cognitive skills. Although the mental efficiency method did not lead to superior test results, it showed training benefits in every study. Furthermore, students are capable to use learner control of learning task selection effectively as shown in Study 2, where the students who trained with learner control exhibited superior performance on a reaction time test. Whereas students seem able to deal with the given control, Studies 3 and 4 indicate that self-ratings should be used with caution. Because these students were novice learners with the FMS, it is conceivable that the novelty of the task at hand disabled their ability to judge their own performance. Of all students in these two studies, only 33% of the students were able to estimate their performance accurately.

Limitations to research

The findings of this dissertation have several implications for the use of personalized instruction in general and the mental efficiency method in particular. While personalized instruction can be beneficial, the research in this dissertation also points out what might have limited possible effects of the training methods. The limitations addressed here are the complexity of the training tasks and the test tasks, the history of training tasks, the role of motivation, and the number of factors in the efficiency formula.

All studies were conducted in a laboratory setting in which novice students were presented with a familiarization training for either ATC or FMS. The level of complexity of the overall material was downsized in order to avoid overloading the participants. For the ATC tasks studies (Chapters 3 and 4), the most important aspects and features in ATC were maintained in a whole-task approach of the training. For the FMS studies (Chapter 5), we focused on the FMS while excluding most of the other cockpit devices. However, this part-task approach was sufficient for our purpose to familiarize the participants with the FMS.

When taking the results of all studies into account, questions arise why the overall performance during both training and test phases seems higher than in comparative studies. Although additional analyses revealed no ceiling or floor effects, it might be that the range of complexity used in our studies was too limited. While most participants needed some time to adjust to the training material at first, they learned to use it rather quickly and kept a steady learning progress. Setbacks during training did not occur very frequently and if they occurred, they only did so in the minority of participants who attained the lowest overall performance in their respective experimental condition.

Not only on learning tasks of the highest complexity levels but even on the most difficult transfer tasks, overall performance remained high.

Furthermore, in all studies the transfer tasks were unable to differentiate between the experimental conditions. The only test effect was revealed by the reaction time test used in the second ATC study (Chapter 4), which showed beneficial effects for personalized task selection conditions, particularly for the learner control condition. The same pattern can be observed in the mental effort scores. The participants indicate moderate levels of invested mental effort in the ATC studies and even lower levels of invested mental effort in the FMS studies. These patterns imply that the overall complexity of the materials used might have been too low and suggest that possibly larger differences in performance and mental effort could have been found with more complex materials. Overall, the participants attained a slightly lower test performance than training performance, but the relatively high test performance scores suggest that they might have been able to execute even more complex tasks.

A further aspect that might have attributed to the limited effects of the training methods might be found in the efficiency method. Originally, the efficiency method was developed to estimate the efficiency of experimental conditions and not to be used as a determinant for dynamic task selection. To use it for this purpose, the relation between performance and mental effort (i.e., efficiency) is estimated for each learning task based on the performance and mental effort scores of the last executed task. The optimization of the learning process might have been limited due to the fact that the efficiency method does not take the history of previous learning tasks and associated performance and mental effort scores into account.

Also, since the original efficiency formula takes only performance and mental effort into account, it is insensitive to other important factors like motivation. However, an indication of the learner's motivation might be found in the relationship between performance and mental effort. While a student who attains a low performance score but yet invests a high amount of mental effort is seen as low efficient according to the efficiency formula, the invested mental effort might also indicate that the student is highly motivated.

When comparing the invested mental effort across all studies, it seems that most students invest only moderate mental effort during training. While initially students are certainly challenged in the training to learn the basic ATC and FMS skills, once they have acquired these skills the challenge becomes less. As a consequence, motivation might decrease when they feel that they are not really challenged anymore.

Furthermore, the original efficiency formula presumes that the other aspects in experimental conditions, like training time, do not differ between the conditions. Indeed, the two ATC studies (Chapters 3 and 4) seem to suggest that the formula is not useful when a third factor differs between the experimental conditions. In the first ATC study (Chapter 3), total training time proved to be an important factor influencing the efficiency of the experimental

conditions. Similarly, the performance on the reaction time test in the second study (Chapter 4) showed that the inclusion of a third factor can lead to a more insightful view on the relation between costs (invested mental effort, number of training tasks, time on task) and benefits (performance, transfer) of training.

Finally, the number of factors in the efficiency formula used for dynamic task selection during training can differ from the number of factors used to determine the efficiency of the experimental conditions after the experiment is completed. In the first ATC study (Chapter 3), total training time was used to determine the efficiency of the experimental conditions after completion of the experiment. However, total training time was not used as a factor for dynamic task selection during training simply because total training time can only be scored after a participant has completed the training. Likewise, the second ATC study (Chapter 4) used performance on the reaction time test as a third factor to determine the efficiency of the conditions but this third factor could not be used for dynamic task selection as well.

Implications and further research

Automation should be used carefully in training programs since it is not a goal in itself but a tool to support the acquisition of skills quickly and efficiently. Because novice learners are easily overloaded with the complexity of an extensive work environment of an Air Traffic Controller or a pilot, it might be good for them to start training with a simplified and less automated training environment. When novice learners have acquired the basic skills, they can be presented with more complex training programs, which increasingly resemble the complete work environment in all its complexity.

In contrast to previous research, the studies in this dissertation have shown that students seem to be able to use learner control efficiently. Students who are given control over the learning tasks and their respective complexity level are able to create an effective training sequence. As long as the level of given control does not overload the students, they can shape their own training sequence. Results have shown that the participants in the learner control conditions usually completed training faster than any other experimental condition while exhibiting mostly equal and sometimes even superior test performance. Further exploration of the level of given learner control, and of how to adapt the amount of control to the growing expertise of the learners during training, represents a promising line of future research.

The use of self-rating should be handled with much caution as the FMS studies (Chapter 5) of this dissertation have shown. While students are able to select an appropriate learning task in terms of complexity, the capacity of estimating the quality of one's own performance is lacking in most students (see also Bjork, 1999; Tousignant & DesMarchais, 2002). Since the students in the

FMS studies were novice learners, it is conceivable that the novelty of the task at hand disabled their ability to judge their own performance. While 66% of all students overestimated their performance, only 33% of the students were able to estimate their performance accurately. For future research it would be interesting to investigate to what extent more advanced students are able to use self-assessment. The 'expertise reversal effect' (for an overview, see Kalyuga, Ayres, Chandler, & Sweller, 2003) shows that instructional materials should be adjusted to the level of learner expertise. The elaborated instructional materials that are helpful at the start of a training program might become redundant when the student has attained a higher level of expertise. Not only might such more advanced students be able to deal with higher levels of learner control but they might also be capable to use self-assessment more accurately than the novice learners in our studies.

Also, in combination with self-assessment, the use of peer-assessment in novice students might lead to interesting effects. Research has shown that peer-assessment positively influences the students' view on learning and assessment, improves learning satisfaction, and enhances clarity of the learning criteria (e.g., Sluijsmans, 2002; Sluijsmans, Moerkerke, Dochy, & van Merriënboer, 2001). Furthermore, by learning to assess their peers, the students reflect more on their own performance (e.g., Anderson & Freiberg, 1995; Gentle, 1994; Longhurst & Norton, 1997; Sobral, 1997) and the awareness of the quality of their own performance improves (e.g., Anderson & Freiberg, 1995; Gentle, 1994; Sluijsmans, 2002). More advanced students who have used peer-assessment in early training phases might also be more capable in rating their own performance in a later training phase.

All studies in this dissertation based task selection on the performance and its associated mental effort of the previous task. However, a student might not pay enough attention or not be very motivated during a specific task, which will have severe consequences in the task selection as used in this dissertation. Such fluctuations will be handled more properly with a student model which not only shows the exact history of performance and mental effort on every learning task a student has completed, but also enables task selection to be based on more solid grounds. Incorporating the history of all learning tasks in the mental efficiency method, and so suppressing the effect of the last task, seems promising for future research.

Furthermore, exploration of an advanced efficiency formula might also prove to be successful. As shown in the first two studies, adding a third factor enabled the formula to take more important aspects into account which leads to a more complete picture of efficiency. Also, the aspect of motivation is interesting to investigate in relation to the efficiency formula. The relationship between performance and mental effort might also be an indication of the learner's motivation. For instance, a student who invests a high amount of

mental effort yet attains a low performance score will be classified as low efficient by the efficiency formula. However, the high amount of invested mental effort might also indicate that the student is highly motivated.

Paas, Tuovinen, van Merriënboer, and Darabi (in press) have recently proposed a task involvement formula that is derived from the original efficiency formula. Future research should take these comments into account in order to examine an improved efficiency method.

Final remarks

The current dissertation can be seen as a first attempt to investigate the possibilities, benefits, and limitations of personalized training methods that are based on an extensive instructional design model such as the 4C/ID model (van Merriënboer, 1997; van Merriënboer, Clark, & de Croock, 2002). Like Chapter 2 shows, the development of personalized training methods has been very diverse and often fragmentary. To use an extensive instructional design model as the basis for training development and to adapt the actual training to the needs of the individual learner is something that has started only recently. Also, the additional use of the concept of cognitive load in the process of dynamic task selection is not to be found in many studies. The research in this dissertation uses the 4C/ID-model (van Merriënboer, 1997; van Merriënboer, et al., 2002) as a starting point. This model offers a training design that presents students with a predetermined order and complexity of learning tasks in such a way that their cognitive capacity is optimally used. While it originally started out as a non-dynamic instructional design model it allows further improvement of the training efficiency by incorporating personalized and adaptive training trajectories.

Though the studies in this dissertation have not delivered indisputable support for the claim that personalized training methods are more effective, they have shown that personalized instruction can have beneficial effects for the individual learner. While some questions are left unanswered and new ones have arisen, this dissertation gives various leads and clues on how to proceed with the investigation of personalized training methods.

References

Anderson, J. B., & Freiberg, H. J. (1995). Using self assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly, 22*, 77-91.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: The MIT Press.

Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort, and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Gentle, C. R. (1994). Thesys: An expert system for assessing undergraduate projects. In M. Thomas, T. Sechrest, & N. Estes (Eds.), *Deciding our future: Technological imperatives for education* (pp. 1158-1160). Austin, TX: The University of Texas.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation, 23*, 319-330.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. (in press). A motivational perspective on the relation between mental effort and performance: Optimizing learners' involvement in instructional conditions. *Educational Technology, Research and Development*.

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & van Merriënboer, J. J. G. (1994a). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology, 86*, 122-133.

Paas, F., & van Merriënboer, J. J. G. (1994b). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371.

Sluijsmans, D. M. A. (2002). *Student involvement in assessment: The training of peer assessment skills*. Doctoral dissertation. Maastricht, the Netherlands: Datawyse.

Sluijsmans D. M. A., Moerkerke, G., Dochy, F., & van Merriënboer, J. J. G. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation, 27*, 153-173.

Sobral, D. T. (1997). Improving learning skills: A self-help group approach. *Higher Education, 33*, 39-50.

Tousignant, M., & DesMarchais, J. E. (2002). Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: A correlation study. *Advances in Health Sciences Education, 7*, 19-27.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology, Research and Development, 50*, 39-64.

Summary

It is becoming more important than ever to educate students in such a way that they can deal with rapid changes in their work and daily life. Ongoing technological developments, especially in domains like aviation and chemical industry in which mistakes can lead to dangerous situations and high costs, ask for continuous adaptation and lifelong learning. Because the available training time is often limited, training should be used as efficiently as possible. Therefore, education must be available whenever it is needed. To cope with these new demands, the educational research field shows two main tendencies: (a) an increasing use of real-life tasks as the driving force for learning, and (b) an increasing flexibility of educational programs as indicated by terms such as "just-in-time-learning" and "education-on-demand".

This dissertation combines both tendencies. A sequence of meaningful learning tasks is used as the backbone of a training program, and the learning task sequence is conceived as a dynamic entity where each next task can be selected in such a way that it best suits the needs of an individual learner. The most optimal learning task sequence is produced by an algorithm for dynamic task selection, which chooses a next learning task by using indications of the student's progress, such as performance scores and costs which are related to attaining these scores (e.g., mental effort, training time, number of learning tasks).

The main research question of this dissertation is how dynamic task selection can be used to optimize training programs, the learning process, and transfer performance. More specific research questions focus on the different types of information that are required to effectively use dynamic task selection and on the role of the trainees themselves in this task selection process. For example, do performance measures contain sufficient information for dynamic task selection or are supplementary measures, such as the amount of mental effort that a student has to invest, also important to take into account for task selection? And, is it possible and desirable to let trainees fulfill an active role in the process of task selection?

In order to be able to address these research questions, Chapter 2 presents a comparison of approaches to learning task selection that have been used throughout the last three decades in training programs for complex cognitive skills. This comparison shows an important development from static part-task selection to dynamic whole-task selection of learning tasks. Four approaches are identified: (a) a static part-task approach, (b) a static whole-task approach, (c) a dynamic part-task approach, and (d) a dynamic whole-task approach. These four approaches are compared in terms of their flexibility and adaptability to the needs of the individual trainee during training. Furthermore,

they are compared to investigate for what complex cognitive skills they might or might not be useful. From this comparison it follows that the *static part-task approaches* can be used well for training procedural tasks with a low organization between part-tasks, like for instance aircraft maintenance. *Dynamic part-task approaches* can be used for tasks that are too complex to practice in a whole-task format. For instance, when learning to drive a car, trainees usually start with learning to steer the car before learning to drive the car in a more integrated whole-task fashion. The static *whole-task approaches* can be used for a training in which the tasks need to be performed in a specific order, like learning to diagnose a clinical case. Lastly, the *dynamic whole-task approaches* can be used for a wide range of training programs to learn complex skills because of their highly flexible and adaptive nature, for instance aviation and the military. If it is possible to find a version of a whole task that is simple enough to start the training with, then these approaches offer the possibility to give trainees a good impression of the professional tasks that they can expect. Furthermore, the incorporation of static approaches in a dynamic approach can yield larger benefits in developing a more effective and efficient training.

Based on the comparison in Chapter 2, Chapter 3 investigates the differential effects of four task selection methods on training efficiency (i.e., training time) and transfer performance in a computer-based training program for Air Traffic Control (ATC). A non-dynamic condition, in which the learning tasks are presented to the participants in a fixed, predetermined sequence, is compared to three dynamic conditions. The three dynamic conditions select learning tasks on the basis of the learner's performance, mental effort, or mental efficiency (i.e., a combination of performance and mental effort). The results support the hypothesis that adapting training to the individual needs of the student makes training more efficient. However, no evidence is found to support the hypothesis that task selection based on mental efficiency leads to more efficient training and higher transfer performance than selection based on either performance or mental effort alone. Despite this result, the mental efficiency condition shows more benefits for the training process than the other dynamic conditions.

Chapter 4 investigates the effects of two personalized training methods on training efficiency and transfer test performance in the same computer-based ATC training program. In one personalized condition (i.e., mental efficiency), task selection is based on mental efficiency which combines performance and invested mental effort; in the other personalized condition (i.e., learner control), the students are free to select the complexity level of the next learning task. Furthermore, participants in both personalized conditions are matched to "yoked" participants in two control conditions. That is, each individualized training sequence of a participant in the mental efficiency condition or the learner control condition is also presented to a participant in the corresponding

yoked control condition. Each yoked participant is presented with the training sequence of *someone else*; hence no personalization occurs in the yoked conditions. In accordance with Study 1 (Chapter 3), the results support the hypothesis that adapting the learning task complexity by means of learner control or mental efficiency makes training more efficient for the individual student. No evidence is found to support the hypothesis that task selection based on personalized efficiency would lead to more efficient training and higher transfer performance than task selection by the students themselves (i.e., learner control). While the mental efficiency condition is effective in terms of learning benefits, the high costs (i.e., mental effort) resulting from these learning benefits cause a low efficiency. In contrast, the learner control condition is not effective in terms of learning benefits, but proves to be an efficient method due to low costs in terms of mental effort.

Chapter 5 presents two closely related empirical studies. The first study examines the effects of three task selection methods on training efficiency and test performance in a computer-based training program for programming a Flight Management System (FMS). A non-dynamic condition, in which the learning tasks are presented to the participants in a fixed, predetermined sequence, is compared to two dynamic conditions, in which the learning tasks are either selected by the participants themselves (i.e., learner control) or by a task selection algorithm. In contrast to the previous studies, in which only self-ratings of mental effort were used, the algorithm in this study uses both self-ratings of mental effort and performance. In agreement with Studies 1 and 2, the results show that the dynamic conditions have beneficial effects on the number of training tasks and the amount of time needed to complete the training, but do not yield higher test performance than the non-dynamic condition. Furthermore, the data suggest that most participants systematically overestimate their performance. Therefore, the role of self-ratings is further investigated in a second study in which the non-dynamic fixed condition, with a smaller amount of learning tasks than in the first study, is again compared to a mental efficiency condition in which students assess their own performance and mental effort. An important finding of this final study is that good self-raters select more appropriate learning tasks and reach higher test performance than bad self-raters.

The final Chapter of this dissertation, Chapter 6, contains a general discussion of the theoretical framework and the empirical studies. Based on the combined results of the studies it is concluded that personalized instruction can have beneficial effects on the training of complex cognitive skills. Although the mental efficiency method did not lead to superior test results, it showed training benefits in every reported study. Furthermore, students are capable to select their own learning tasks (learner control), as shown in Study 2. The results of Studies 3 and 4 put this effect into perspective and show that self-

assessment should be used with caution. Because these students were novice learners in regard to the Flight Management System (FMS), it is conceivable that the novelty of the task at hand disabled their ability to judge their own performance. This has several implications for the use of personalized task selection and the mental efficiency method. While personalized task selection can be beneficial for training, the research in this dissertation also points out what might have limited possible effects of the training methods. The limitations addressed here are (a) the complexity of the training and test tasks, (b) the history of training tasks, (c) the role of motivation, and (d) the number of factors in the efficiency formula. It is concluded that automation of task selection should be used carefully in training programs since it is not a goal in itself but a tool to support the acquisition of skills quickly and efficiently.

For future research it is interesting to investigate to what extent more advanced students are able to use self-assessment. Also, in combination with self-assessment, the use of peer-assessment by novice students might lead to interesting effects on the selection of learning tasks. Incorporating a history of learning tasks in the mental efficiency method, instead of using only the last learning task as in the current studies, also seems promising for future research. Finally, the use of an elaborated efficiency formula might prove to be successful for dynamic task selection in education.

Samenvatting

Het wordt steeds belangrijker om studenten zo op te leiden dat zij kunnen omgaan met snelle veranderingen in hun werkomgeving en dagelijkse leven. Vooral de voortschrijdende technologische ontwikkelingen in domeinen zoals de luchtvaart en chemische industrie waarin fouten tot gevaarlijke situaties en hoge kosten kunnen leiden, vragen om voortdurende aanpassingen en het levenslang leren van mensen. Omdat de beschikbare trainingstijd vaak beperkt is dient training zo efficiënt mogelijk ingezet worden. Daartoe moet onderwijs altijd beschikbaar zijn op het moment dat het nodig is. Om aan deze nieuwe eisen tegemoet te kunnen komen zijn er binnen het onderwijsonderzoek en de onderwijspraktijk twee tendensen waarneembaar: (a) een toename van het gebruik van levensechte taken als de motor voor betekenisvol leren, en (b) een toenemende flexibiliteit van onderwijskundige programma's in termen van 'just-in-time-learning' en 'education-on-demand'.

Deze dissertatie combineert beide tendensen. Er wordt gebruik gemaakt van een opeenvolging van betekenisvolle leertaken als de ruggengraat van een trainingsprogramma. De opeenvolging van leertaken wordt als een dynamische entiteit beschouwd door elke volgende leertaak zo te kiezen dat deze het beste aansluit bij de behoeften van een individuele student. Daarbij wordt de optimale opeenvolging van leertaken bepaald door een algoritme voor dynamische taakselectie, dat een beslissing neemt over de volgende leertaak op basis van indicatoren van de voortgang van de student. Indicatoren zoals prestatiescores en kosten die gerelateerd zijn aan het behalen van deze scores (bijvoorbeeld mentale inspanning, trainingstijd, aantal benodigde leertaken etc.).

De belangrijkste onderzoeksvraag in deze dissertatie richt zich op de manier waarop dynamische taakselectie gebruikt kan worden om trainingsprogramma's, het leerproces, en de transferprestatie te optimaliseren. Meer specifieke vragen hebben betrekking op de verschillende soorten informatie die nodig zijn om dynamische taakselectie effectief te kunnen gebruiken, alsook de rol van de student zelf in dit taakselectieproces. Geven prestatiemetingen bijvoorbeeld voldoende informatie voor dynamische taakselectie, of zijn aanvullende metingen, zoals de hoeveelheid mentale inspanning die de student moet leveren, ook belangrijk om te gebruiken ten behoeve van taakselectie? En is het mogelijk en wenselijk om studenten zelf een actieve rol te laten spelen in het taakselectieproces?

Om deze onderzoeksvragen te kunnen beantwoorden presenteert Hoofdstuk 2 een overzicht van methoden van leertaakselectie, welke gedurende de laatste drie decennia in trainingsprogramma's voor complexe cognitieve vaardigheden gebruikt zijn. Dit overzicht laat een belangrijke ontwikkeling zien

van een statische opeenvolging van *deel*taken naar dynamische selectie van hele leertaken. Er worden vier methoden geïdentificeerd: (a) statische deeltaakselectie, (b) statische hele-taakselectie, (c) dynamische deeltaakselectie, en (d) dynamische hele-taakselectie. Deze vier methoden worden vergeleken op hun flexibiliteit en adaptatievermogen in relatie tot de behoeften van de individuele student tijdens de training. Verder wordt vergeleken voor welke complexe cognitieve vaardigheden de verschillende taakselectiemethoden wel of niet geschikt zijn. Deze vergelijking laat zien dat de *statische deeltaakmethoden* vooral geschikt zijn voor het trainen van procedurele vaardigheden met weinig organisatie tussen deeltaken, zoals 'aircraft maintenance'. *Dynamische deeltaakmethoden* zijn vooral geschikt voor het trainen van complexe vaardigheden die de student in het begin van de training nog niet als hele taak aankan. Denk bijvoorbeeld aan leren autorijden, waarbij de lerende in eerste instantie alleen het stuur hoeft te bedienen, alvorens met meer geïntegreerde hele-taakoefening verder te gaan. *Statische hele-taakmethoden* zijn vooral geschikt voor het trainen van taken waarbij stappen in een bepaalde volgorde uitgevoerd moeten worden, zoals het stellen van een medische diagnose. De *dynamische hele-taakmethoden*, ten slotte, zijn geschikt voor een breed scala aan trainingsprogramma's voor het leren van complexe cognitieve vaardigheden vanwege hun flexibiliteit en adaptatievermogen. Hierbij valt bijvoorbeeld te denken aan de luchtvaart en het leger. Als het lukt om een versie van de taak te vinden die eenvoudig genoeg is om de training mee te beginnen bieden deze selectiemethoden de mogelijkheid om de lerende al snel een goede indruk te geven van de hele taak. Tevens kan het opnemen van statische methoden in een dynamische methode tot een meer effectieve en efficiënte training leiden.

Op basis van het overzicht in Hoofdstuk 2 worden in Hoofdstuk 3 de effecten van vier taakselectiemethoden op trainingsefficiëntie (in termen van trainingstijd) en transferprestatie onderzocht in een computergestuurd trainingsprogramma voor het leiden van luchtverkeer ("Air Traffic Control"). Een niet-dynamische conditie, waarin de leertaken in een van tevoren vastgestelde opeenvolging aan de proefpersonen worden aangeboden, wordt vergeleken met drie dynamische condities. In de drie dynamische condities worden leertaken geselecteerd op basis van prestatie, mentale inspanning, of mentale efficiëntie (d.w.z., een combinatie van prestatie en mentale inspanning). De resultaten ondersteunen de hypothese dat het aanpassen van de training aan de behoeften van de individuele student de training meer efficiënt maakt. Er wordt echter geen ondersteuning gevonden voor de hypothese dat taakselectie op basis van mentale efficiëntie tot meer efficiënte training en betere transferprestatie leidt dan taakselectie op basis van alleen prestatie of alleen mentale inspanning. Desondanks vertoont de mentale efficiëntie conditie meer positieve effecten op het verloop van de training dan de andere dynamische condities.

Hoofdstuk 4 onderzoekt de effecten van twee gepersonaliseerde trainingsmethoden op trainingsefficiëntie en transferprestatie in, opnieuw, een computergestuurd trainingsprogramma voor het leiden van luchtverkeer. In één gepersonaliseerde conditie ("mental efficiency") worden leertaken geselecteerd op basis van een mentale efficiëntiemaat die prestatiescores en mentale inspanningsscores combineert; in de andere gepersonaliseerde conditie ("learner control") zijn de studenten vrij in het kiezen van de moeilijkheidsgraad van de volgende leertaken. Verder zijn de proefpersonen in beide gepersonaliseerde condities gekoppeld aan "yoked" proefpersonen in twee controlecondities. Dit betekent dat elke individuele trainingsvolgorde van een proefpersoon in de "mental efficiency" of de "learner control" conditie ook wordt aangeboden aan een proefpersoon in de corresponderende "yoked control" conditie. In de yoked control condities krijgen de proefpersonen dus een trainingssequentie van *iemand anders*; en er vindt dus geen personalisatie plaats. In overeenstemming met Studie 1 (Hoofdstuk 3) geven de resultaten steun aan de hypothese dat het aanpassen van de moeilijkheid van de leertaken door middel van "learner control" of "mental efficiency" tot meer efficiënte training voor de individuele student leidt. Er wordt echter geen bewijs gevonden voor de hypothese dat taakselectie op basis van mentale efficiëntie tot meer efficiënte training en betere transferprestatie leidt dan taakselectie door de student zelf ("learner control"). De "mental efficiency" conditie is effectief in termen van leeruitkomsten, maar blijkt niet erg efficiënt te zijn omdat de kosten in termen van mentale inspanning hoog zijn. De "learner control" conditie blijkt niet effectief in termen van leeruitkomsten, maar is wel efficiënt omdat de geïnvesteerde mentale inspanning laag is.

Hoofdstuk 5 rapporteert twee gerelateerde empirische studies. De eerste studie onderzoekt de effecten van drie taakselectiemethoden op trainingsefficiëntie en testprestatie in een computergestuurd trainingsprogramma voor het leren programmeren van een vluchtbeheerssysteem ("Flight Management System"). Een niet-dynamische conditie, waarin de leertaken in een van tevoren vastgestelde volgorde aan de proefpersonen worden aangeboden, wordt vergeleken met twee dynamische condities. In de twee dynamische condities worden, net als in Studie 2 (Hoofdstuk 4), leertaken geselecteerd door de proefpersonen zelf ("learner control") óf door een taakselectiealgoritme in het computergestuurde programma dat in tegenstelling tot de voorgaande studies niet alleen gebruikt maakt van "self-ratings" voor mentale inspanning maar ook voor prestatie. In overeenstemming met Studies 1 en 2 wordt aangetoond dat de dynamische condities betere effecten bewerkstelligen op het aantal benodigde trainingstaken en de trainingstijd maar geen hogere testprestatie laten zien dan de niet-dynamische conditie. Verder suggereren de gegevens dat veel proefpersonen hun prestatie systematisch overschatten. Daarom wordt de rol

van "self-ratings" verder onderzocht in een tweede studie waarin de niet-dynamische conditie, met minder leertaken dan in de eerste studie, wederom wordt vergeleken met een mentale efficiëntie conditie waarin studenten hun eigen prestatie en mentale inspanning beoordelen. Een belangrijke bevinding van deze laatste studie is dat goede "self-raters" meer geschikte leertaken selecteren en een hogere testprestatie behalen dan slechte "self-raters".

Het laatste hoofdstuk van deze dissertatie, Hoofdstuk 6, bevat een algemene discussie van het theoretische kader en de empirische studies. Op basis van de gecombineerde resultaten van de verschillende studies wordt geconcludeerd dat gepersonaliseerde taakselectie tot betere training voor complexe cognitieve vaardigheden kan leiden. Hoewel de mentale efficiëntiemethode niet tot superieure testresultaten heeft geleid, vertoont deze methode in elke gerapporteerde studie wel positieve effecten op de training. Verder zijn studenten goed in staat om zelf leertaken te selecteren ("learner control"), zoals blijkt uit Studie 2. De resultaten van Studies 3 en 4 relativeren deze bevinding enigszins en tonen aan dat "self-asssessment" in ieder geval niet zondermeer gebruikt kan worden. Omdat de studenten beginners waren met betrekking tot het vluchtbeheerssysteem ("Flight Management System") is het denkbaar dat de nieuwigheid van de taak hun vermogen om de eigen prestatie te beoordelen verstoorde. Dit heeft enkele implicaties voor het gebruik van gepersonaliseerde taakselectie en de mentale efficiëntiemethode. Hoewel gepersonaliseerde taakselectie tot meer efficiënte training kan leiden duidt het onderzoek in deze dissertatie ook aan wat de oorzaken kunnen zijn voor het ontbreken van positieve effecten van dynamische taakselectiemethoden. Mogelijke oorzaken betreffen (a) moeilijkheidsgraad van de trainings- en testtaken, (b) de leertaakgeschiedenis, (c) de rol van motivatie, en (d) het aantal factoren in de efficiëntieformule. Automatisering van taakselectie dient in trainingsprogramma's met mate gebruikt te worden omdat het geen doel op zichzelf is, maar slechts een gereedschap om het snel en efficiënt leren van vaardigheden te ondersteunen.

Voor toekomstig onderzoek is het interessant om te bestuderen in hoeverre meer gevorderde studenten in staat zijn om hun eigen prestatie te beoordelen. Tevens kan "peer-assessment" in combinatie met "self-assessment" voor beginnende studenten tot interessante bevindingen bij het zelf selecteren van leertaken leiden. Het meenemen van de geschiedenis van alle uitgevoerde leertaken in de mentale efficiëntiemethode, in plaats van alleen de direct voorafgaande leertaak zoals in de huidige studies, lijkt ook veelbelovend voor toekomstig onderzoek. Ten slotte kan het gebruik van een uitgebreide efficiëntieformule succesvol blijken voor dynamische taakselectie in het onderwijs.

Curriculum Vitae

Ron Salden was born on the 22nd of December, 1976, in Geleen, the Netherlands. In 1996 he finished secondary education at the 'Groenewald Proosdijveld' high school in Stein. From 1996 until 2000 he studied Psychology at the University of Maastricht. He specialized in Cognitive Psychology and completed this study program with a thesis on a study which examined the effects of an unfamiliar score and randomization on music expertise. After his graduation he started working on a PhD project at the Educational Technology Expertise Center (OTEC) of the Open University of the Netherlands (OUNL). The project entailed the examination of personalized training methods for complex cognitive skills in the aviation domain.

ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the following universities: University of Twente, University of Groningen, Maastricht University, University of Amsterdam, Utrecht University, Open University, Leiden University, Wageningen University, Technical University of Eindhoven, and Free University (and formerly University of Nijmegen and University of Tilburg). More than a hundred dissertations have been published in this series. The most recent are listed below.

86. Fox, G.J.A. (07-09-2001). *Multilevel IRT: A Bayesian Perspective on Estimating Parameters and Testing Statistical Hypotheses*. Enschede: University of Twente.
87. Brok, P.J. den (21-09-2001). *Teaching and Student Outcomes: A Study on Teachers' Thoughts and Actions from an Interpersonal and a Learning Activities Perspective*. Utrecht: Utrecht University.
88. McKenney, S.E. (12-10-2001). *Computer-Based Support for Science Education Materials Developers in Africa*. Enschede: University of Twente.
89. Oosterheert, I.E. (15-11-2001). *How Student Teachers Learn: A Psychological Perspective on Knowledge Construction in Learning to Teach*. Groningen: University of Groningen.
90. Stoyanov, S.T. (16-11-2001). *Mapping in the Educational and Training Design*. Enschede: University of Twente.
91. Eysink, T.H.S. (07-12-2001). *Signs for Logic Teaching*. Enschede: University of Twente.
92. Geijsel, F.P. (13-12-2001). *Schools and Innovations: Conditions Fostering the Implementation of Educational Innovations*. Nijmegen: University of Nijmegen.
93. Fukkink, R.G. (02-02-2002). *Instructing Primary School Children in Deriving Word Meaning from Written Context*. Amsterdam: University of Amsterdam.
94. Vosse, A.J.M. (10-04-2002). *Evaluatie van het Tutorleren in Nederland: Een onderzoek naar de cognitieve en sociaal-emotionele effecten van een tutorprogramma voor rekenen-wiskunde in het basisonderwijs*. Amsterdam: University of Amsterdam.
95. Jager, B. de (25-04-2002). *Teaching Reading Comprehension: The Effects of Direct Instruction and Cognitive Apprenticeship on Comprehension Skills and Metacognition*. Groningen: University of Groningen.

96. Veldhuis-Diermanse, E.A. (26-04-2002). *CSCLearning? Participation, Learning Activities and Knowledge Construction in Computer-Supported Collaborative Learning in Higher Education*. Wageningen: Wageningen University.
97. Crombach, M.J. (02-05-2002). *Sixth and Seventh Graders' Appraisal Processes of Curricular Tasks: Development of a Model to Explain Students' Willingness to Invest Effort*. Nijmegen: University of Nijmegen.
98. Kruiter, J.H. (13-05-2002). *Groningen Community Schools: Influence on Child Behaviour Problems and Education at Home*. Groningen: University of Groningen.
99. Braaksma, M.A.H. (21-05-2002). *Observational Learning in Argumentative Writing*. Amsterdam: University of Amsterdam.
100. Lankhuijzen, E.S.K. (24-05-2002). *Learning in Self-Managed Management Career: The Relation between Managers' HRD-Patterns, Psychological Career Contracts and Mobility Perspectives*. Utrecht: Utrecht University.
101. Gellevij, M.R.M. (06-06-2002). *Visuals in Instruction: Functions of Screen Captures in Software Manuals*. Enschede: University of Twente.
102. Vos, F.P. (26-06-2002). *Like an Ocean Liner Changing Course: The Grade 8 Mathematics Curriculum in the Netherlands, 1995-2000*. Enschede: University of Twente.
103. Sluijsmans, D.M.A. (28-06-2002). *Student Involvement in Assessment: The Training of Peer Assessment Skills*. Heerlen: Open University.
104. Tabbers, H.K. (13-09-2002). *The Modality of Text in Multimedia Instructions: Refining the Design Guidelines*. Heerlen: Open University.
105. Verhasselt, E. (14-11-2002). *Literacy Rules: Flanders and the Netherlands in the International Adult Literacy Survey*. Groningen: University of Groningen.
106. Beekhoven, S. (17-12-2002). *A Fair Chance of Succeeding: Study Careers in Dutch Higher Education*. Amsterdam: University of Amsterdam.
107. Veermans, K.H. (09-01-2003). *Intelligent Support for Discovery Learning: Using Opportunistic Learner Modeling and Heuristics to Support Simulation Based Discovery Learning*. Enschede: University of Twente.
108. Tjepkema, S. (10-01-2003). *The Learning Infrastructure of Self-Managing Work Teams*. Enschede: University of Twente.
109. Snellings, P.J.F. (22-01-2003). *Fluency in Second Language Writing: The Effects of Enhanced Speed of Lexical Retrieval*. Amsterdam: University of Amsterdam.
110. Klatter, E.B. (23-01-2003). *Development of Learning Conceptions during the Transition from Primary to Secondary Education*. Nijmegen: University of Nijmegen.
111. Jellema, F.A. (21-02-2003). *Measuring Training Effects: the Potential of 360-degree Feedback*. Enschede: University of Twente.

112. Broekkamp, H.H. (25-02-2003). *Task Demands and Test Expectations: Theory and Empirical Research on Students' Preparation for a Teacher-made Test*. Amsterdam: University of Amsterdam.
113. Odenthal, L.E. (21-03-2003). *Op zoek naar balans: Een onderzoek naar een methode ter ondersteuning van curriculumvernieuwing door docenten*. Enschede: University of Twente.
114. Kuijpers, M.A.C.T. (21-03-2003). *Loopbaanontwikkeling: onderzoek naar 'Competenties'*. Enschede: University of Twente.
115. Jepma, IJ. (01-07-2003). *De schoolloopbaan van risico-leerlingen in het primair onderwijs*. Amsterdam: University of Amsterdam.
116. Sotaridona, L.S. (05-09-2003). *Statistical Methods for the Detection of Answer Copying on Achievement Tests*. Enschede: University of Twente.
117. Kester, L. (05-09-2003). *Timing of Information Presentation and the Acquisition of Complex Skills*. Heerlen: Open University.
118. Waterreus, J.M. (05-09-2003). *Lessons in Teacher Pay: Studies on Incentives and the Labor Market for Teachers*. Amsterdam: University of Amsterdam.
119. Toolsema, B. (23-10-2003). *Werken met competenties*. Enschede: University of Twente.
120. Taks, M.M.M.A. (20-11-2003). *Zelfsturing in leerpraktijken: Een curriculumonderzoek naar nieuwe rollen van studenten en docenten in de lerarenopleiding*. Enschede: University of Twente.
121. Driessen, C.M.M. (21-11-2003). *Analyzing Textbook Tasks and the Professional Development of Foreign Language Teachers*. Utrecht: Utrecht University.
122. Hubers, S.T.T. (24-11-2003). *Individuele leertheorieën en het leren onderzoeken in de tweede fase*. Eindhoven: Technical University of Eindhoven.
123. Sun, H. (04-12-2003). *National Contexts and Effective School Improvement: An Exploratory Study in Eight European Countries*. Groningen: University of Groningen.
124. Bruinsma, M. (09-12-2003). *Effectiveness of Higher Education: Factors that Determine Outcomes of University Education*. Groningen: University of Groningen.
125. Veneman, H. (01-07-2004). *Het gewicht van De Rugzak: Evaluatie van het beleid voor leerlinggebonden financiering*. Groningen: University of Groningen.
126. Annevelink, E. (27-08-2004). *Class Size: Linking Teaching and Learning*. Enschede: University of Twente.
127. Vries, B. de (15-10-2004). *Opportunities for Reflection: E-mail and the Web in the Primary Classroom*. Enschede: University of Twente.
128. Veenhoven, J. (05-11-2004). *Begeleiden en beoordelen van leerlingonderzoek: Een interventiestudie naar het leren ontwerpen van onderzoek in de tweede fase van het voortgezet onderwijs*. Utrecht: Utrecht University.

129. Strijbos, J.W. (12-11-2004). *The Effect of Roles on Computer-Supported Collaborative Learning*. Heerlen: Open University.
130. Hamstra, D.G. (22-11-2004). *Gewoon en Anders: Integratie van leerlingen met beperkingen in het regulier onderwijs in Almere*. Groningen: University of Groningen.
131. Lubbers, M.J. (09-12-2004). *The Social Fabric of the Classroom: Peer Relations in Secondary Education*. Groningen: University of Groningen.
132. Nijman, D.J.J.M. (10-12-2004). *Supporting Transfer of Training: Effects of the Supervisor*. Enschede: University of Twente.
133. Dewiyanti, S. (25-02-2005). *Learning Together: A Positive Experience. The Effect of Reflection on Group Processes in an Asynchronous Computer-Supported Collaborative Learning Environment.* Heerlen: Open University.
134. Stoof, A. (04-03-2005). *Tools for the Identification and Description of Competencies.* Heerlen: Open University.
135. Salden, R.J.C.M. (22-04-2005). *Dynamic Task Selection in Aviation Training.* Heerlen: Open University.