Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer.

Open Universiteit
www.ou.nl

Running head: PERSONALISED ADAPTIVE TASK SELECTION IN AIR TRAFFIC

CONTROL

Personalised Adaptive Task Selection in Air Traffic Control:

Effects on Training Efficiency and Transfer

Ron J. C. M. Salden, Fred Paas,

Jeroen J. G. van Merriënboer

Open University of the Netherlands

Correspondence concerning this article should be addressed to Ron J.C.M. Salden, Open

University of the Netherlands, Educational Technology Expertise Center (OTEC), P.O. Box

2960, NL-6401 DL Heerlen, The Netherlands. E-mail may be sent to ron.salden@ou.nl

phone: 0031-45-5762398, fax: 0031-45-5762907

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 personalised conditions were compared with two corresponding yoked control conditions. The hypothesis that personalised adaptive task selection leads to more efficient training than non-adaptive task selection was partially confirmed. However, the hypothesis that adaptive task selection based on personalised efficiency leads to more efficient training than adaptive task selection based on personalised preference was not supported. The results are discussed and suggestions are given for future research.

 Keywords: personalised adaptive task selection, personalised training, mental efficiency, learner control

Personalised Adaptive Task Selection in Air Traffic Control:

Effects on Training Efficiency and Transfer

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. An efficient training offers trainees an optimal

learning environment in which they can acquire skills quickly and adequately while also learning

how to apply these skills flexibly to new situations, i.e. transfer.

A theoretical comparison by Salden, Paas, and van Merriënboer (2004) showed that

training methods for complex cognitive skills have evolved in three important ways during the

last three decades. The *first* change marks a shift in focus from non-adaptive to adaptive

methods. Although both methods take prior knowledge into account in the development of the

training, non-adaptive approaches can only determine the sequence of learning tasks prior to the

start of the training. Yet the adaptive approaches also have the possibility to make adjustments in

the task sequence during training. The *second* distinction 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), which the trainee is not able to practice 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 meaningful parts to train.

The *third* change shows a shift from group-based to personalised methods. Whereas group-based

training can be very useful in terms of allocated time and resources, the intricate nature of complex cognitive skills makes different demands for each individual student.

Following from these three changes, especially the use of personalised adaptive whole-task selection is believed to be strongly related to increased training efficiency (Salden, Paas, Broers, & van Merriënboer, 2004). The group-based non-adaptive 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 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-adaptive methods exhibit a high dropout rate of ATC students.

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

As the role of the computer developed significantly, training programs started to become computer-based enabling trainers the possibility to make adjustment 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 personalised adaptive 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 operationalise 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., personalised 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, a frequent finding has been that especially low-ability learners do not make good use of 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' (e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 1998, 2001) indicates that a trainee's 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 personalised 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 are used. According to Paas, Tuovinen, Tabbers and van Gerven (2003) mental effort can be considered to reflect the cognitive load as it refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task. 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 adaptive selection of learning tasks.

When trainees achieve the same performance scores, mental effort might be able to reveal differences 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. However, most people would agree that the next learning task should be assigned to the first person, whereas the second person could be given a more difficult learning task next. 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 personalised task selection, and consequently training efficiency can be improved by taking these measures into account. When using both measures for personalised adaptive task selection, the presented learning tasks fit well to the cognitive schemata a student has acquired. The individual capacity of a student is well taken into account, hence leading to high training efficiency.

A first indication that the use of a combined performance and mental effort score (i.e., personalised efficiency) might make personalised training more efficient was found in two recent studies of Camp, Paas, Rikers, and van Merriënboer (2001), and Salden et al. (2004). Both

studies compared four methods of system-controlled 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 adaptively, based on either performance or

mental effort, or the combination of both (i.e., mental efficiency). Results showed that

personalised adaptive task selection leads to more efficient training than group-based non-

adaptive task selection. Although personalised adaptive task selection based on mental efficiency

did not lead to more efficient training and better transfer than personalised adaptive task

selection based on performance or mental effort alone, it did show several training benefits in

terms of jump size, number of training tasks and training time.

The combined measure of performance and mental effort, as used in the aforementioned

studies (Camp et al., 2001; 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' behaviour 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 a high mental effort is least efficient.

Hypotheses

In the current study we compared non adaptive task selection conditions to two

personalised adaptive conditions, in which task selection was either based on mental efficiency

(system-controlled) or based on personal preference (learner-controlled). Based on the aforementioned benefits it is hypothesized that personalised adaptive task selection leads to more efficient training and better transfer task performance than non-adaptive task selection. Since both adaptive conditions personalise the task selection either by using the performance and the associated mental effort of the individual student or by giving learners a level of control that is fairly compatible with their abilities, these conditions are expected to lead to superior training efficiency.

The second hypothesis states that adaptive task selection based on personalised efficiency would lead to more efficient training and better transfer than selection based on personalised preference. While the efficiency condition adjusts to the needs of the individual learners, learners in the personalised preference condition have to select training tasks on their own. 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.

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 (about $ 23) for their participation.

Materials and Procedure

Learners were first presented with an introduction to the domain and second, received training using an ATC-trainer where they had to execute 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.

In the terms of Crooks and Klein (1996), learners are given the option to bypass instructional material. To avoid confusion due to the yoked control condition, learner control will be called personalised preference in the remainder of this article.

Introduction

At the beginning of the experiment, the participants were given a Microsoft® PowerPoint® presentation, which contained 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 (see below for a more detailed description of the complexity levels). 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 familiarised with the way the aircraft reacts 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 adaptive ATC-situations on a radar

screen, in which a number of possible conflicts had been built in. Participants were required to

guide the 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 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. The interface is depicted in Figure 1.

INSERT FIGURE 1 ABOUT HERE

Learning Task Complexity

Prior to training, the 77 learning tasks were divided into 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 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. In consultation with professional Air Traffic Controllers from EUROCONTROL 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 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 lets participants proceed to the transfer tasks after being presented with 20 learning

tasks. This amount of tasks ensures sufficient variation over the ten complexity levels. The second outcome states that a participant has completed the training when s/he has achieved a score that meets the preset performance and mental effort criteria on two learning tasks of the highest complexity. In the last possible outcome, participants complete 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 can 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) personalised efficiency condition, (2) yoked efficiency condition, (3) personalised preference condition, and (4) yoked preference condition.

(1) Personalised efficiency condition. The selection of learning tasks in the personalised efficiency condition was based on the combination of performance and mental effort measures. 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. Together with performance, these subjective ratings of mental effort were used for determining the complexity level of the next learning task.

In order to be able to determine the level of performance achieved by a participant when working on a learning task of a particular complexity level, we used the performance data of a previous experiment. Since the participant population was similar to a previous experiment (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 personalised 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 (2004) experiment were used to formulate scoring tables for all complexity levels from which a mean performance score could be derived.

INSERT TABLE 1 ABOUT HERE

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 a particular complexity level is depicted in Table 1. To obtain the maximum score (100%) for each performance variable of this task, a participant had to attain four gate hits by giving no more than four commands, spending no time outside the airways, and maintaining sufficient separation between aircraft.

INSERT TABLE 2 ABOUT HERE

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 2); 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 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.

INSERT TABLE 3 ABOUT HERE

Filling in both performance and mental effort scores in the selection table (Table 3) 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. 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. The mean performance scores in this table correspond to the scores in

Table 2. Assume that a participant has used 6.67 commands (75%), 96 out of airways (25%), 22.67 no separation (50%) and 2 gate hits (50%) for the accomplishment of a specific learning task of complexity level 4. Averaging this performance results in a mean performance score for the task of 50%. Table 2 indicates that this corresponds to a mean performance score of 3. This mean performance score together with the mental effort rating provided by the participant is used in Table 3 to determine whether the complexity of the next learning task should be increased, decreased or remain the same. For example, assume that the participant had provided a mental effort rating of 1, then the complexity of the next learning task would be increased by two levels, whereas a mental effort rating of 4 would result in a decrease of complexity by one level for the next learning task.

INSERT FIGURE 2 ABOUT HERE

----

(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 personalised efficiency condition were assigned to participants in the corresponding yoked condition. Each training sequence was allocated only once to one participant in the yoked efficiency condition.

(3) Personalised preference condition. After completing the pre-training a window popped up in which the participants in the personalised 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. The participants were unfamiliar with the task attributes determining 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. Consequently, the maximum jump size in this condition was 8.

(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 personalised preference condition were assigned to 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. Five dependent variables were registered, namely 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 absolute jump size between complexity levels. Moreover performance and mental effort (subjective rating scales) during training were assessed. Analyses of variance (ANOVAs) and planned comparisons will be used to analyse the data.

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 the mean RT on conflict identification and the number of correct conflict identifications were determined.

Secondly, participants were required to solve ten transfer tasks that were structurally different from the learning tasks in such a way that new aircraft frequently appeared in the

interface, some of the aircraft had a different speed than other aircraft while all aircraft had the same speed in the learning tasks, and the number of aircraft that had to be directed to their appropriate landing spot was larger than in any of the learning tasks. Performance and mental effort on these transfer tasks were registered. The transfer tasks were scored on the same four variables as the training tasks, namely time outside airway, no separation, number of commands, and gate hits. As for training performance, each variable was scored between 0-100% of highest obtainable performance and subsequently these scores were added and divided by the number of performance variables to get one overall transfer performance score.

Lastly, 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. (2004), in which efficiency is calculated on the basis of three – rather than two - 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):

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

In this formula, E = mental efficiency, ME = mental effort during test, RT = reaction time performance, and TT = test performance.

<div align="center">Results</div>

<div align="center">*Training Phase*</div>

<div align="center">--</div>

<div align="center">INSERT TABLE 4 ABOUT HERE</div>

Training Effects

No effects (all *Fs* < 1) 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. 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 ($t < 1$) was found between the

personalised conditions and the yoked conditions. With regard to the second hypothesis, planned

comparisons showed that the personalised efficiency condition reached a higher mean

complexity level (t (56) = 2.19, $p < .05$) than the personalised 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 ($t < 1$) was found between the personalised conditions and the yoked

conditions. However, planned comparisons did support the second hypothesis because the

personalised efficiency condition made larger jumps in complexity levels (t (56) = 2.69, $p < .05$)

than the personalised preference condition.

Performance and mental effort during training. 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, η^2 = .13). With regard to the first hypothesis, the mean performance score of the personalised efficiency and personalised 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 personalised efficiency condition and its corresponding yoked condition (t (56) = 2.84, p < .01). Furthermore, planned comparison did support the second hypothesis because the personalised efficiency condition attained a higher performance (t (56) = 2.44, p < .05) than the personalised preference condition. Means and standard deviations are provided in Table 4.

INSERT TABLE 5 ABOUT HERE

Transfer Test Phase

Reaction time test. No effect (F < 1) for the independent variable Method of Task Selection was found on the variable mean RT on conflict identification. 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, η^2 = .31. Planned comparisons confirmed the first hypothesis since the mean number of correct conflict identifications (M = 29.60, SD = 5.58) of the personalised efficiency condition and personalised 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 personalised preference condition attained more conflict identifications (t (56) = -3.58, $p < .01$) than the personalised efficiency condition.

Transfer tasks. When analysing 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).

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 ($t < 1$) was found between the personalised conditions and the yoked conditions. In contrast with the second hypothesis, planned comparisons showed that the personalised efficiency condition was less efficient (t (56) = -3.00, $p < .01$) than the personalised preference condition. The means and standard deviations are provided in Table 5.

Discussion

The main hypothesis of this study that personalised adaptive task selection leads to more efficient training and better transfer performance than non-adaptive task selection was partially confirmed. While a strong trend showed that the mean training performance score of both personalised conditions was higher than the mean performance score of their corresponding yoked conditions, further analysis revealed that this was only caused by the difference between the personalised efficiency condition and its corresponding yoked condition. With regard to test performance, the mean number of correct RT conflict identifications of both personalised conditions was higher than the mean number of conflict identifications of their yoked conditions. This finding is in agreement with other studies showing superiority of adaptive selection of

learning tasks over non-adaptive selection (e.g., Camp et al., 2001; Kalyuga & Sweller, 2005; Salden et al., 2004; Van Merriënboer et al., 2002).

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

The small amount of differences between the personalised conditions and their associated yoked conditions might be attributed to the pre-training. While its purpose was to familiarise 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.

As another explanation, it might be that the personalised conditions lead to individual training sequences that were also useful for most participants in the yoked conditions. Despite the fact that the sequences were not personalised for the yoked participants, the progress their

linked participants made 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 also choose more or less the same training sequence. If our participants had had varying expertise levels, larger differences in the training sequences would have been expected, yielding more benefits of the personalized conditions over the yoked conditions.

It is less easy to explain the unexpected finding that the personalised efficiency condition had positive effects on the mean complexity level, jump size, and performance during training but that, in contrast, the personalised preference condition showed more correct conflict identifications and higher transfer test efficiency. Possibly, this finding is related to the fact that task selection in the personalised efficiency condition was only based on their overall performance, and not on the particular type of conflict the participants had trouble with. In one complexity level different types of conflict could 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, in contrast to participants in the personalised preference condition, 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. Consequently, participants in the personalised efficiency condition had less opportunity to learn from their mistakes than participants in the personalised preference condition. This difference might have resulted in more performance-oriented processing in the personalised efficiency condition (i.e., directed towards fast improvement of overall performance) and more knowledge-oriented

processing in the personalised preference condition (i.e., directed towards remediation of

particular misunderstandings), which could be an explanation for the training benefits of the

efficiency condition and the transfer benefits of the preference condition. Future research could

focus on fostering transfer performance in personalised efficiency conditions by adjusting the

task selection process in such a way that more knowledge-oriented processing is stimulated.

Future research should attempt to gain a better insight into the cognitive processes

associated with the different task selection methods. Only then, it will be possible to explain the

differential effects of these methods on training and transfer performance, and, consequently, to

manipulate their effectiveness. To uncover cognitive processes, process-tracing methods, such as

verbal reporting ('thinking aloud'), eye tracking, and decision analysis, could be used. The

information elicited with these methods would allow for making inferences about the cognitive

processes or knowledge underlying performance. Another, more indirect, way to create a better

understanding of the underlying cognitive processes would be to administer a delayed transfer

test some time (e.g., two weeks) after the training is given. Delayed transfer testing can be

considered a more sensitive way to reveal differences in the quality of the constructed cognitive

schemata than direct transfer testing.

In contrast to various studies on learner control, the present study showed that students

are capable of using the given control when the training situation is well constructed and avoids

overloading their cognitive system. When even novices are able to deal with a certain degree of

freedom, it would be interesting to investigate to what extent more experienced participants

might be able to use learner control. With students that already have acquired a certain level of

expertise, one could experiment with increasing the degrees of freedom. With a higher expertise

level, it would be expected that the student would benefit most from a large amount of given freedom.

The use of the extended 3D efficiency formula 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 could be inserted into the 3D formula. This led to a realistic efficiency score that represents a complete view on the transfer phase which was not possible using the 2D version of the formula as shown by additional analyses (F < 1). 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 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 personalised efficiency would lead to more efficient training and higher transfer than selection based on personalised preference. While the personalised efficiency condition showed several training benefits, it did not prove to lead to higher transfer performance nor was this condition efficient. In contrast, the personalised preference condition showed only minor training benefits, yet it did prove to 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 adaptive methods, it has been shown that it could not compete with the personalised 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., & 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.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. Educational Technology Research and Development, 53, 83-93.

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., & 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., 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.

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

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 task selection approaches. *Computers in Human Behavior*.

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.

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., de Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills, 84*, 784-786.

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.

Footnotes

[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).

[a] The slight differences on the five training phase variables between the personalised 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.

Table 1

Example of a Scoring Table for the Performance Variables of a Learning Task of a Particular

Complexity Level

	Score				
	0%	25%	50%	75%	100%
Number of 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 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

Table 4

Overview of Training Results

| | Method of Task Selection | | | | | | | |
| | Efficiency | | Yoked efficiency | | Preference | | Yoked preference | |
Dependent variables [a]	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

Table 5

Overview of Transfer Test Results

Dependent variables	Efficiency		Yoked efficiency		Preference		Yoked preference	
	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
Efficiency	-0.43	0.62	-0.42	0.89	0.60	1.13	0.25	1.04

The column header "Method of Task Selection" spans the four method columns (Efficiency, Yoked efficiency, Preference, Yoked preference).

Figure Captions

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

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

the next learning task.