BEYOND THE SIMULATOR

INSTRUCTION FOR HIGH-PERFORMANCE TASKS

Het in dit proefschrift beschreven onderzoek werd uitgevoerd bij TNO Technische Menskunde in Soesterberg, binnen de context van het Interuniversitair Centrum voor Onderwijsonderzoek.

The research reported in this dissertation was carried out at TNO Human Factors, Soesterberg, the Netherlands, in the context of the Interuniversity Center for Educational Research

Printed by Print Partners Ipskamp
Cover design: Martijn van Emmerik / Walter van Dijk

ISBN: 90-365-2068-1

BEYOND THE SIMULATOR

INSTRUCTION FOR HIGH-PERFORMANCE TASKS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. F. A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op woensdag 22 september 2004 om 13.15 uur

door

Martijn Louis van Emmerik

geboren op 2 juni 1972
te Haarlem

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. A.J.M. de Jong
en
prof.dr. J. Pieters

Voorwoord

Het voltooien van een proefschrift vereist naast een gezonde dosis intelligentie een enorme hoeveelheid doorzettingsvermogen. Wat dat betreft mag ik me gelukkig prijzen dat beide eigenschappen rijkelijk vertegenwoordigd waren in mijn naaste omgeving gedurende de afgelopen jaren. Mijn collega’s van de UTwente en TNO-TM verdienen dan ook wel een bedankje.

Als ‘AIO op afstand’ heb ik niet zo direct kunnen deelnemen aan de dagelijkse gang van zaken op de UTwente. Toch heb ik altijd betrokkenheid gevoeld van secretaresses, kamergenoten, en andere collega’s als ik weer eens langskwam voor een verjaardagviering, cursus, ProIST meeting, een werkbespreking, of een combinatie van deze activiteiten.

Ook op TNO-TM heb ik het in de afgelopen jaren naar mijn zin gehad. Wat betreft werk, maar zeker ook op het sociale vlak. Waar vind je zo’n open sfeer als binnen ‘Training en Opleiding’, zulke ‘erudiete’ gesprekken tijdens de koffiepauze, en kun je kiezen uit volleyballen, tafeltennissen of hardlopen tussen de middag.

Een aantal mensen verdient een persoonlijk bedankje. Als ik dit proefschrift met een raket vergelijk moet ik Wouter van Joolingen en John van Rooij bedanken voor de lancering. Zij gaven me brandstof in de vorm van stapels literatuur, en samen discussieerden we over mogelijke reisbestemmingen. Al snel echter, verdwenen zij uit het zicht en had ik vooral contact met het ‘Mission Control Center’ te Enschede bemand door de constante factor Jules Pieters. Zonder op de voorgrond te treden las hij steeds al mijn schrijfsels zowel inhoudelijk als op punten en komma’s door en gaf daarbij waardevolle commentaren.

Ton de Jong heeft scherp gelet op de terugkeer naar de aarde en de landing. Hij verdient credits voor doorzettingsvermogen ("wetenschap stopt niet om vijf uur") maar zeker ook voor zijn inhoudelijke commentaar. Als ik met mijn ‘Zeeuwse schrijfstijl’ ("geen woord teveel hoor!") mijn conclusies te veel in de ruimte liet hangen bracht hij me altijd weer ‘in orbit’.

Na het vertrek van John bij TNO heb ik regelmatig gediscussieerd met Herke Schuffel die vast een deel van zijn inbreng herkent in dit boek. Diens vertrek was voor Alma Schaafstal aanleiding om het estafettestokje over te nemen (en wie zal het verbazen dat we het soms ook nog over hardlopen hadden). Daarnaast ben ik veel dank verschuldigd aan Hans Korteling die een belangrijke bijdrage aan mijn wetenschappelijke ontwikkeling geleverd heeft vanaf het moment dat ik als stagiair bij TNO binnenkwam. Ook de discussies met mijn andere collega's van Training en Opleiding zijn vaak inspirerend geweest.

Zonder de technische ondersteuning van Wytze Hoekstra, Ko Stoel en Ingmar Stel had ik echter geen experiment kunnen draaien. Hierbij moet ook Robin Sip die de databases voor de simulator leverde genoemd worden, en Simone Brouwer die alle proefpersonen streng doch rechtvaardig van instructie en feedback voorzag. Koos Wolff en Walter van Dijk, om nog een paar collega's te noemen, verleenden ondersteuning bij de vormgeving van dit proefschrift.

Ten slotte moet ik natuurlijk Hélène bedanken die me altijd wist aan te sporen wanneer mijn eigen motivatie even niet toereikend was, en Sanne omdat ze nog voldoende vragen heeft waar ik geen antwoord op heb.

Martijn van Emmerik
Almere, augustus 2004

Contents

Chapter 1

General Introduction

Abstract
High-performance tasks are difficult because of their complexity and the fact that they are time critical. Because they often have to be executed in a dangerous or hostile environment they require an extensive amount of training before they can be executed safely. From an operational as well as a didactical point of view, simulation offers many advantages to train these tasks. The didactical possibilities of simulation, however, don't seem to be exploited on a large scale in training currently. The present research focuses on tutoring to provide insight in the utility of augmented cues. Central question is if deviating from reality (by enhancing the salience of critical cues or adding new cues during tutoring) can increase effectiveness and efficiency of the training process.

1.1 High-performance tasks and simulation

Some tasks require more training than others do. There is more to this observation than simply saying that these tasks are probably 'more difficult' than tasks that require little or no training. Tasks can be difficult for different reasons, e.g., they may have to be performed in a limited time frame, they are composed of many subtasks, or maybe they have to be performed under dangerous conditions. For some tasks even, this all applies. In those cases, it is appropriate to speak of 'high-performance tasks'.

High-performance tasks are defined as complex, time-critical tasks where the operator is in the primary control loop of the system (Van Rooij, 1994; Schneider, 1985). An example is piloting a (combat) helicopter. The time-critical aspect of this task derives from the fact that the to-be-controlled system is dynamic and requires continuous adjustments because of its inherent instability (Hart, 1988) Furthermore, it operates in a dynamic and often dangerous or hostile environment. The complexity of flying a helicopter resides in the number of, variety of, and interactions between multiple requisite skills. Apart from the perceptual-motor skills, time-sharing with (subsidiary) procedural and cognitive skills for communication, orientation, and interpretation of instrument readings is required.

One of the training characteristics of high-performance skills is that many people fail to develop proficiency so that selection prior to training is sometimes required. After selection, the training duration required to reach an operational level of performance usually is considerable. Typically, there are large differences between novice, advanced, and expert operators, not merely with respect to the speed and accuracy of performance, i.e. quantitatively, but also with respect to the use of different strategies, i.e., qualitatively (Chapman, Underwood, & Roberts, 2002; Underwood, Chapman, Bowden, & Crundall, 2002; Lesgold et al., 1988).

A variety of high-performance tasks ranging from air-traffic control (ATC) to different kinds of vehicle control tasks (flying an aircraft, driving a car, etc…) is trained by means of simulation because this offers certain advantages for training[1].

Learning in a simulation environment proceeds through interaction with (and manipulation of) a model. A model is a representation of (a part of) the real world in mathematical equations. Models are by definition a simplification of reality and as such, they can be used very well for training and instruction purposes. According to Alessi and Trollip (2001), this feature helps learners to construct their own mental representation of the simulated phenomenon by directing focus on the task-relevant aspects of the model. For this purpose, an interface is required to ‘translate’ the manipulations of the trainee (operator) into mathematical values the model can ‘understand’, process, and report the outcomes back to the learner via that same interface. In simulations of high-performance tasks (vehicle control tasks in particular) the ensemble of mathematical model and interface is called a simulator (Van den Bosch & Riemersma, 2000). This term emphasizes the presence (and

[1] Chapter 3 goes into this in more detail

importance) of a physical representation of the controls (mock up). This makes sense because the handling of the controls is a main aspect of vehicle control. For other types of tasks, where it is more important to understand how the simulated system (process) works the interface is generally ‘computer based’. Because of the absence of a physical mock-up, these systems are usually called simulations - emphasizing the process in which the application is used. The difference between simulator and simulation thus is mainly one of terminology in which the emphasis shifts from modeling a physical system to modeling a process.

1.2 Challenges

For high-performance tasks, it is particularly interesting that the use of a simulator for training creates opportunities to control a training program to an extent that would be impossible in reality. This implies that simulator training could be didactically more advanced than training on the real system, at least theoretically. In practice, however, the didactic possibilities of a simulator as a training device are often limited because of methodological problems and conflicting interests of the parties involved.

Generally, there is a lack of knowledge about training and instructional factors in relation to simulators and high-performance tasks. Consequently, simulator training is often approached in exactly the same way as traditional training. Because only a part of the simulators' potential is used in these cases, the observed training efficiency of many simulators is limited[2] (Polzella, 1983; Verstegen, Barnard, & Van Rooij, 1999).

The accumulation of knowledge in the field is slow because of methodological problems. Research that compares instruction in a simulator with training on the real system will be difficult to interpret because of inherent differences between the two training systems. The experimental setup often makes it impossible to determine to what extent between-group differences should be attributed to the training program or to the characteristics of the system configuration that was used for training. Comparing the effectiveness of two different training programs therefore can only be done if they were both administered on the same system configuration (whether that is the real system or a representation of it). This kind of study is rarely seen.

Apart from that, educational researchers and simulator designers appear to have different interests. The latter have directed most of their efforts to the optimization of technical aspects of simulators in an attempt to create the best approximation of the real system as it is assumed that this will result in the optimal transfer of training. Although not explicitly stated, the instructional aspects seem to be taken for granted in this ‘hardware-focused’ approach[3].

[2] Surprisingly enough, users and instructors often appear to be very much satisfied with the possibilities their training devices offer (Verstegen et al, 1999).
[3] Frequently, training requirements are not considered until after the simulator has been bought (Farmer, Van Rooij, & Riemersma, 1999).

Within the educational sciences on the other hand, simulation is mainly applied as a tool to explore (interactively) the relations that exist between variables in scholarly tasks (e.g. physics). These simulations function as experimentation environments for (high-) school children and intend to stimulate development of cognitive and procedural skills by means of discovery learning (e.g., De Jong, 1997; De Jong et al., 1998; Van Joolingen, 1993). For training of perceptual-motor skills, which are often associated with high-performance tasks, however, this form of simulation is not applicable. The absence of a mock-up precludes the practicing of perceptual and motor aspects whereas these are the very aspects that should be similar between training task and transfer task (Schmidt, 1982). The differences between both approaches have stood in the way of the development of an integrative account of high-performance training using simulators.

1.3 The present research

Insight into factors that determine the effectiveness of simulator-based training is indispensable in designing, procuring, and using training simulators optimally. The present research aims at the development of knowledge with regard to the design of effective and efficient simulator training for high-performance tasks. The focus of the experimental part of this thesis is on tutoring i.e., support during execution of training activities.

Merrill, Reiser, Ranney, and Trafton (1992) describe tutoring as a process in which a trainee learns by doing. The role of the tutor in this process is to provide some kind of assistance to prevent the trainee from floundering too much. It is not necessary for the tutor to intervene immediately whenever an error is made. Only when the trainee errs in such a way that he or she will take too much time to get back on the track, help is provided. It may be expected that the role of the instructor diminishes (or at least changes) with the growing skills of the trainee. Metaphorically speaking, the instructional support can be seen as a process of scaffolding and fading. These illustrative terms have also been adopted in the training and instruction literature (e.g., Patrick, 1992; Smith & Ragan, 1999).

The tutoring process is central to apprenticeship teaching (an instructional method typically used for learning high-performance tasks) in which a master or expert directly supervises a pupil performing a task (Brown, 1989). Initially the pupil watches the master perform but eventually he has to complete the task alone. The expert monitors the performance of the trainee and selects events (as they occur) as examples for learning. During this process, the learner receives immediate feedback and the master gradually withdraws the amount of support.

The opportunities to provide feedback in the synthetic environment a simulator provides are enhanced in comparison to reality. It is possible to adapt naturally occurring cues to increase their salience and so optimize their value in the learning process. In fact, artificial cues can even be created to support performance and relieve operator workload. This is called augmentation of (virtual) reality (O'Shea, Cook, & Young, 1999). Augmented cues can be a very potent instructional factor. They increase the salience of certain aspects in the environment guiding the trainee to perceive those cues that are critical to (or relevant for) correct performance. If these augmented cues are activated only when the trainee is in need of help they function somewhat like training wheels on a bicycle: they provide a safe envelope within one can practice driving on two wheels.

Intuitively, the concept of augmentation is very appealing. Nevertheless, there is no overwhelming empirical support for it. Although there are several accounts that support the concept of augmentation (e.g., Lintern & Koonce, 1992; Lintern & Roscoe, 1982) in some cases trainees become dependent on the cues, which are obviously absent in reality.

In the present research it was hypothesized that a simulator using advanced training aspects (such as augmented cues) would be more effective than one that did not make use of these features. Note that the latter variant would be more realistic as no such cues are available in reality. In other words, this hypothesis suggests that deviation from reality during training can enhance transfer. This idea was used to formulate the following research question:

- How should tutoring in a simulator take place to increase effectiveness and efficiency of the training process? In particular, is it better to stay to reality as close as possible or can it be beneficial to deviate from reality during training.

This question is too general to answer. Therefore, three (sub) questions have been derived from it:

- How do instructions in a high-performance task relate to the learning process?
- How is the efficiency of training in a simulator affected by two different approaches to tutoring: one 'traditional' method based on verbal instruction, and one 'experimental' method based on augmented cueing and feedback?
- Tutoring is a process that revolves around interaction between tutor and trainee. To what extent then do trainee characteristics interact with the efficiency of these two variants of tutoring?

After describing the theoretical background for these questions from the available literature, the second part of this thesis contains the description of three experiments that were conducted to provide an answer to the research questions.

The first experiment was mainly explorative and served to establish a baseline of performance with 'standard' instruction (i.e., without advanced features). Additionally, an attempt was made to predict the instructions from observed behaviors.

A second experiment was done to investigate the effects of the two different ways of presenting instructions and finally, the third experiment added the role of trainee characteristics to this aspect.

In order to get around the major source of interpretation difficulties mentioned in the previous paragraph, in the present research a simulator-to-simulator design was applied rather than making a comparison between a simulator and reality.

A consequence of the simulator-to-simulator design is that experimental results will not lead to statements about transfer to any real-world task. This was taken for granted as the main interest was in the effects of the instructional manipulations on itself. Although the transfer question is a very interesting one, no attempts were made to answer it here.

1.4 Outline / Research approach

In this thesis, car driving has been chosen as a generic example of a high-performance task. The reasons for this choice are theoretical as well as pragmatical. Car driving is quite a difficult task, in particular from the perspective of a novice driver. As explained in chapter 4, driving involves many different skills. The devils' advocate might argue that driving a car is by far not as difficult as flying a combat helicopter (see example in paragraph 1.1). However, it incorporates all the necessary aspects of Schneiders' (1985) definition: it is complex, time critical, and the operator is in the primary control loop.

A more pragmatical reason is that the population of potential car drivers is much larger than that of helicopter pilots which makes it easier to find subjects for experiments.

After this general introduction chapter, chapters 2 and 3 provide the theoretical basis for the research questions leading to the three experiments that were conducted in a driving simulator environment. Chapter 2 provides the theoretical background of instruction in relation to simulation. It provides definitions of the used concepts and an overview of the relevant literature with regard to high-performance tasks and training. In Chapter 3, it is claimed that simulators are not used to their full capacities. Verstegen, Barnard, and Van Rooij (1999) showed that the practical considerations and technical (hardware) developments have received most attention in the development of training simulators. Particularly in those tasks that seem to have relatively few possibilities to benefit from practical advantages (such as car driving), the possibilities of the didactic approach present a promising step to new and better approaches to simulator training.

By specifically taking advantage of the didactical opportunities of simulators, a substantial increase in training effectiveness and - or efficiency can be accomplished. To achieve this, relevant knowledge from instructional science needs to be integrated with expertise in the field of simulator development. Whenever this knowledge is insufficient, not applicable, or unavailable, further research should be instigated.

Chapter 4 serves as a pivot between the theoretical chapters of this thesis and the descriptions of the three experiments. It further focuses the ideas from high-performance tasks in general to car driving in specific. These theoretical considerations meet the 'practice' in the paragraphs about low cost driving simulation where a description of the experimental environment is provided

The first experiment is described in Chapter 5. Initially focusing on the instruction and feedback during tutoring, no use of simulator specific features was made yet. Instead, the experiment tried to find the 'rules' that guided the provision of instruction in the simulator under different conditions during the training process.

The findings were used to design two forms of instruction (one rooted in 'traditional' verbal instruction, the other -simulator specific- based on the concept of augmented cues) to be used in the two follow-up experiments. These are reported in Chapter 6. The first of these two experiments basically compares the two forms of instruction and feedback. In the final experiment the factor aptitude is introduced as it is expected to mediate the effectiveness of instruction and feedback.

This dissertation concludes with Chapter 7 that summarizes the findings from theory as well as the experiments and sketches the consequences for instruction in high-performance tasks in the discussion.

Chapter 2

Training for High-performance Tasks

Abstract
From a training point of view it is possible to look at tasks, skills and their development in many different ways. Currently there is no single approved view that guarantees success in the design of training. In this chapter several approaches are discussed, each with its specific strengths. Yet, all models have difficulties predicting the duration of the learning process and explaining the mediating factors. Especially because these factors are expected to differ depending on the level of detail an instructional designer takes. To provide insight in these factors, a choice is made to focus the current research on the learning process at the lowest (event) level and on the support that is provided during the accompanying training activities (tutoring).

2.1 Introduction

Theoretical developments in relation to training of high-performance tasks can be seen from two main perspectives: the traditional simulator design perspective (e.g., Allen, Hays, & Buffardi, 1980; Hogue et al. 1999; Zeltzer & Pioch, 1996) and the training perspective (e.g., Alessi, 1988; Lintern, Roscoe, Koonce, & Segal, 1990; Roscoe, 1991, Verstegen & Barnard, 1999). The traditional simulator design viewpoint is very much oriented towards fidelity (see section 3.3). Its adherents emphasize the importance of physical simulator characteristics to facilitate the training process. To put it differently, they consider the characteristics of the hardware and the mathematical models to be the main determinants of the simulator's potential to make training effective and efficient. The advantage of this viewpoint is that specifications for hardware can be excellently quantified. For example, Padmos and Milders, (1992) give an extensive description of the physical factors (such as resolution, luminance, contrast, image complexity, delay, etc…) that affect the quality of computer generated imagery (CGI).

The risk that looms for this approach, however, is the neglect of instructional factors such as training strategies that should be used, necessary training time, sequencing of instructional events, and the instructional support during performance. This insight provides the basis for the present chapter: The instructional value of a particular simulator-configuration may differ depending on the characteristics of the training program that is used -regardless of the technical (hardware and software) specifications of the simulator.

2.2 Tasks and skills

In relation to training in general, two concepts take a central place: 'task' and 'skill'. Being closely related, they are easily (albeit incorrectly) used interchangeably. According to Fleishman and Quaintance (1984), the definitions of the term 'task' that are used in the literature differ on two major dimensions. The first dimension pertains to the breadth of coverage of the definition, i.e., ranging from a single specific performance such as pressing a button under certain conditions to the 'totality of the situation imposed on the performer'. In the latter case, more of a context is provided for the task. Such a context is likely to increase realism because it can provide meaning to otherwise isolated activities. Nevertheless, the interrelationships that undoubtedly exist between sub tasks within an extensive context will complicate task analysis and likewise, development of training.

The second dimension reflects the extent to which a task definition is intrinsic or external. On the one end of this dimension, a task can be defined as a specific performance requirement originating externally from the performer. According to this view, tasks are conceived of as 'given'. On the other hand, it can be argued that a performer will redefine the imposed requirements in terms of personal experiences, past, and needs. Hence, a person’s conception of a task will be subjective. This means that the task definition of the same performance standard will differ for each individual depending on the perceptions a person has of what is being expected from him or her (i.e. determined by intrinsic aspects).

Fleishman and Quaintance (1984) emphasize that no position on any of these dimensions should be preferred a-priori over another. The point is to make the best choice for a specific purpose, whether that is training design, workload assessment, or classification of behavior.

With the above discussion in mind, the definition of tasks used in this thesis is partially adapted from Farmer et al. (1999). They describe a task as any kind of a *performance standard* that is defined independently of the person who executes the task (i.e. extrinsic). Additionally, tasks require goal directed and meaningful activities that are performed in a context. Without context and meaning, an action can support a sub-task at the most.

Now the definition of a task has been formulated, we can contrast it with a skill. A skill refers to the *potential* a person has to perform a particular act. Unlike a task, a skill is dynamic and can change (preferably improve) as a result of a training process (Farmer et al, 1999). In fact, training is essentially aimed at the acquisition and enhancement of skills. A skill, however, is an inferred capability of a person and as such it cannot be observed directly, but has to be deduced from behavior in a test situation -i.e. performance on a task (Patrick, 1992). This notion is essential for the design of training programs because a failure to select the proper exercises (i.e. training tasks) will result in inefficient or (even worse) ineffective skill development.

It will be clear that the main problem for instructional design is how to choose the appropriate task to increase skill level (and thus improve task performance). This is a difficult question because the mapping of skills onto tasks cannot be explained in a simple one to one relationship. A skill can be used to perform different types of tasks and conversely, a single task may be performed using any one of a number of different skills. To make things even more complicated, most real world tasks require a number of different skills anyway. Designing training then becomes a process in which it should be determined which skills should be enhanced at what time.

For training design it is relevant that skills can be conceived of at different levels of complexity. What can be called a high-performance skill from one perspective can also be decomposed into smaller constituents. Such constituent skills (or sub-skills) may be common to many tasks. ‘Lever positioning’ for example is seen in vehicle control but also in process operating. Nevertheless, if many sub-skills need to be integrated into one composite skill, it is likely that the latter will become specifically linked to a single task such as ‘landing an aircraft on a carrier vessel’. In these cases, it will be insufficient to focus on the constituents in isolation for training. The whole task becomes more than the sum of its parts and additional training is required to integrate the constituents.

2.2.1 Analysis of skills and tasks

Within a task analysis, the instructional designer aims to provide understanding of the nature of the desired learning outcome. Therefore, it is considered the most important component process in instructional design (Jonassen, Tessmer, & Hannum, 1999). By means of task analysis, it is possible to unravel the structure of a task. Once the task is divided in manageable units (sub-tasks), it will be possible to determine what skills are required for each sub-task. Then the learning goals can be drawn up after which an instructional strategy has to be selected along with the training tasks to reach these goals (Gagné, Briggs, & Wager, 1992).

The instructional design process would be much helped if a comprehensive classificatory system of human performance were available. Such a system could explicitly links tasks and skills to educational objectives that can be translated systematically and effectively into a curriculum.

A point of criticism for the task analysis approach is that for high-performance tasks, the whole task is considered as more than the sum of its parts. The concurrent processing of different tasks during performance requires the development of additional (organizational / high-performance) skills. Such higher order skills complicate the process of task analysis considerably.

To determine the level of detail that is appropriate to describe skills and tasks for training, a (training) need analysis should be done (Jonassen, Tessmer, & Hannum, 1999). The goal of such an analysis is to identify capabilities of learners, desired skill level, and the discrepancies between those. Two important factors that play a part in the way such an analysis is conducted are the task that requires training (including the environment or conditions of performance) and the characteristics of the trainee (novice vs. expert, intelligence, history, etc.) (Farmer et al, 1999). For example, learning to move a file into a different folder using the Windows Explorer requires skills such as reading, typing, and mouse-handling skills.

One way to perform this task is to press down (once) the left-button of the mouse to select the file, press (and hold) the right-button of the mouse, and drag the file to another location. After releasing the right-button, a pop-up menu appears. The appropriate action [‘move here’] should now be selected from it by moving the cursor over the desired alternative and pressing the left-button again.

For a novice computer user, these actions may be too complex and first some basic mouse handling should be acquired such as pointing to different items on the desktop and clicking them.

Operating that same computer with the goal to find some information on the Internet requires a much richer knowledge of computers (Lazonder, Biemans, & Wopereis, 2000). The important aspects of this task are selection of the best search engine for the question at hand and learning how to use logical operators (in that search engine). Furthermore, the hits that are generated have to be skimmed to see if the search yielded the desired result.

Although a similar level of mouse handling skills is required as in the file-movement task, the emphasis in this second task is completely on other, (predominantly cognitive) skills. The constituent skills (reading, mouse handling, typing, etc.) are considered to be fully mastered. In this example it would be senseless to describe all the mouse actions that should be taken just like a more cognitive approach probably would not work for the 'folder moving' example.

2.2.2 Taxonomies of tasks

The usefulness of a classificatory system of human performance was already mentioned above. In theory, tasks can be divided neatly into categories. However, the fact that a multitude of taxonomies has been proposed (an overview can be found in De Landsheere, 1989) indicates that there is little consensus on what would be the best way to categorize tasks. Of course, different taxonomies are constructed from different points of view (Fleishman and Quaintance, 1984). Their goal can be merely theoretical, (describe the structure and relationships of the constituent objects within a domain systematically, generate hypotheses, order observations, etc.) or utilitarian (define instructional objectives, select a training program for a task, evaluate learning processes). These different approaches to classification then are all valid and useful from their own perspective.

Taxonomies are not only made with different goals in mind but also with focus on different aspects of human performance. With regard to training, the cognitive domain has received most attention hence this is the best developed. Although many task analyses have been done in the psychomotor domain, most of these were not related to training and therefore they have not resulted in useful taxonomies for training. Apart from that, a drawback of any classificatory system is that tasks in the real world often cannot be classified easily because they possess characteristics of multiple categories. Blind application of taxonomies to a real world domain will then result in a rather contrived classification.

Farmer et al. (1999) stress the importance of taxonomies although they contend that most classificatory systems are either too global or not fully applicable to a specific domain. It is clearly impossible (and probably undesirable) to find one generic classificatory system that is able to describe all aspects of human performance. A more serious problem is that human performance is subject to change during training. In other words, the relationship between tasks and skills is not fixed. For these reasons, Farmer et al. (1999) suggest a hybrid approach for training design. Their approach implies that only a crude classification (into cognitive tasks, motor tasks, and procedural tasks) is made after which a task analysis can be conducted to refine these categories with regard to the specific goals of the training designer. In this pragmatic view, taxonomies are not meant to be complete. They are in the first place helpful to encourage instructional designers to think systematically about the different tasks and skills and to consider the effect of training in this light. When design of instruction is a goal (as opposed to the creation of a comprehensive domain description), the existence of so many different taxonomies is actually positive because instructional designers can select the system that best fits their goals.

2.2.2.1 Classifications for high-performance skills

A major problem in designing effective training programs for high-performance skill development is that the knowledge base with respect to training and instruction is not well organized. It simply suffers from a lack of theoretical cohesion (Van Rooij et al., 1995; Van Rooij et al., 1997). Most of the guidelines that are encountered in the literature are formulated rather generally although they have been developed in the context of specific tasks. This renders it difficult to apply them. Another major problem is that most of the research focuses on individual performance and skill acquisition in the context of relatively simple, well-structured tasks. In particular, there is a paucity of research dealing with the integration of skills e.g., (part-task) training of time-sharing skills.

The value of taxonomies in the domain of high-performance skills is that they can provide a framework to divide these tasks into manageable bits. This is helpful for the process of instructional design. By means of classification, it is possible to draw generalizations across events and tasks. Following this, specific performance standards can be defined as a point of reference for training outcome. The development of a categorization of human performance thus helps to connect particular categories of tasks and skills to training techniques.

Gagné, for example, divides human performance (capabilities) and their possible learning outcomes into five categories (see Gagné, 1985; Gagné, Briggs, & Wager, 1992; Smith & Ragan, 1999). Each of these categories is supposed to possess unique characteristics that should be reflected in requirements for training and instruction or, as Gagné states, they differ in the conditions most favorable for learning. These conditions are partly internal (memory) and partly external to the learner. The process of deliberately arranging these external conditions is what Gagné (1985) calls instruction.

Many different ways to categorize tasks and skills have been proposed, three of these are presented below. These are focused on skills, tasks, and their interaction respectively. A point of criticism to make in advance is that such classifications often disregard the fact that skills may develop as a result of training.

Discrete vs. continuous skills
Tasks requiring discrete skills have determined beginnings and endings whereas the start and ending of tasks that depend on continuous skills is more subtle and determined by the performer (Smith & Ragan, 1999). Serial skills can be placed in between; these are actually made up of discrete actions that are performed in an (almost) continuous sequence such as with playing a musical instrument. An example of a continuous skill is steering a car on a winding road. Discrete skills are required for example to shift to second gear or to fasten a bolt.

Self paced vs. force paced tasks
In self-paced activities such as lawn mowing, or changing a tire, the speed of working is determined by the performer (notwithstanding deadlines). Force-paced activities on the other hand strictly dictate the timing of each action that should be taken (e.g. working on an assembly line, or many process operation tasks).

Automatic vs. controlled

A very important distinction (also with regard to instructional design) is formed by the difference between automatic and controlled processes (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The distinction refers to the amount of attentional control that is supposed to be necessary for executing a task. Without elaborating on the controversies that surround the concept of attention (see e.g., Korteling, 1994) it can be said that automatic processing occurs when after sufficient training no (or rather very little) attention is required to perform a task. In these occurrences performance is fast, apparently effortless and it does not interfere with other tasks (provided these tasks do not rely on the same physical or cognitive structures[4]). Automatic processing develops for so called closed skills (Poulton, 1974). Closed skills are used in a predictable environment that cannot actively affect performance, hence the term ‘closed’. Typing is an often-cited example of a skill that can be executed automatically (e.g., Gentner, 1988; Wickens, 1992). Skilled typists, for example, have been reported to carry on a conversation while maintaining an error free performance of 300 to 540 strokes per minute (Gentner, 1988).

Controlled processing on the other hand is the dominant mode of performance in an unpredictable (interactive) environment. Because the environment may have an effect on human action, the associated skills are called open. Even after long periods of training, tasks requiring open skills are executed under conscious control and they require considerable effort and attention. Therefore, performance is relatively slow although it is more flexible than for closed skills (Patrick, 1991).

The related distinction between open- and closed-loop processes (Adams, 1971; Schmidt, 1982) could introduce some confusion because open-loop processes are associated with closed skills and vice versa. The term ‘closed-loop’ namely implies that performance is a continuous process whereby the feedback delivered by a system is (consciously) interpreted against a reference so that corrective actions can be taken. It takes place in a (closed) system that reacts to input and provides feedback just as is the case for open skills. Open skills can thus be influenced by environmental stimuli so that a closed feedback-loop is formed.

Open-loop processes are not subject to events in the environment. They are executed in a preprogrammed sequence just as closed skills. The absence of a feedback loop and the reference of correctness explain the name ‘open-loop’ (Schmidt, 1982).

Although subdivisions in still other categories are possible, the problem with any categorization is that in reality tasks are not executed in isolation (as in a laboratory). Most meaningful tasks (high-performance tasks in particular) require a number of different skills. Driving a car, the task we are investigating here, requires perceptual motor skills to follow the track, and to turn the steering wheel in order to maintain a safe and steady lateral position. Another aspect of driving is shifting gears. This can be considered a procedural task. Nevertheless, a considerable amount of motor skill is required to carefully balance the depressing of clutch pedal and accelerator when

[4] In such cases performance is a matter of efficient switching between tasks rather than dual task performance (i.e. serial instead of parallel execution of two tasks).

driving away in first gear: the driver has to feel the ‘biting point’ of the clutch pedal as it engages. It is likely that with automation of performance (Fitts and Posner, 1967) the perceptual motor components of a task will be executed in a procedure like manner.

Furthermore, car driving requires intellectual skills, and attitudes to interpret traffic signs, and interact with other traffic whereas finding the best route to a destination while integrating information from road maps with other sources of traffic information requires cognitive strategies.

Apart from mastering these (more or less individual) skills, proficient task performance is also about ‘managing’ the amount of attention that should be allocated to each aspect of a complex task at any given moment during performance. It is supposed that a special kind of ‘high-performance skill’ can be developed that enables a person to combine task performance in an optimal way (Schneider, 1985; Wickens, 1992).

2.2.2.2 Classification and stage accounts

Development of skills is often supposed to proceed through (a certain number of) stages. An early stage account of skill development that influenced many others has been postulated by Fitts (1965; Fitts & Posner, 1967). According to Fitts, skills develop through three stages: the cognitive, associative, and autonomous stage.

In the first stage, the trainee has to verbalize the procedures that are part of performance and practice the intellectual component of the skill. During the associative phase, the correct behavioral patterns are established by means of practice and elimination of errors. In the final, autonomous stage, speed and accuracy of performance gradually increase whereas interference from other (concurrent) activities is reduced in a process called automation.

Anderson (1982, 1992) has also developed a stage theory. Although it is focused on cognitive skill acquisition rather than the development of high-performance skills it can easily be mapped onto Fitts' three-phase theory. Anderson distinguishes between the declarative stage, knowledge compilation, and tuning. More explicitly than Fitts, Anderson describes the consequences of his stages as they influence the instructional process. Initially performance relies on facts stored in working memory. By means of two processes, composition and proceduralisation this knowledge is 'compiled' into task-specific procedures. During composition, rules are combined to form new, more direct rules. Proceduralisation is involved with making rules more specific. This is necessary to speed up performance: by creating specific instances of a general rule, working memory demands are decreased and rules can be executed faster. In the final tuning stage, three mechanisms help to improve performance further. Generalization, discrimination, strengthening. By means of these mechanisms the range of applicability of a rule can be broadened or narrowed, or the strength of a rule can be increased or decreased depending on its successful application.

Critics of Anderson (and Fitts) have noted that these theories are descriptive rather than predictive. The duration of each stage for different skills cannot be predicted using these theories. Another point of criticism refers to the strictly sequential nature of these models (Patrick, 1991). It may not always be necessary to pass through a declarative stage before the procedural stage is reached. Some knowledge (implicit knowledge) cannot be verbalized and as such it cannot be transferred from working memory to long-term memory in the way Anderson suggests (Broadbent, 1990; Hayes and Broadbent, 1988).

In a way, the work of Rasmussen (1986) can be related to the theories of Fitts and Anderson. Although not particularly focused on training, the distinction between skill-, rule-, and knowledge-based behavior is similar to the above mentioned stages. For example, Rasmussen describes skill-based behavior as smooth and 'automated', mostly seen in sensory motor performance. Skill based activities are initiated with an intention, yet their performance takes place without further conscious control. In the words of Fitts (1962) this would be called autonomous.

Rule based behavior is, to a certain extent, consciously controlled. Usually, a stored rule or procedure triggers performance. The rule itself is 'released' by a situation that has a certain familiarity. From this experience, a goal directed response could be made. During behavior, other rules can be triggered by cues that may originate from the environment or from the person's behavior itself.

During situations for which no rules for control are available, control of performance is moved up yet another level: knowledge based behavior is characterized by explicit formulation of a goal. Plans to reach this goal are developed (and tested physically or conceptually) by means of functional reasoning.

An additional distinction Rasmussen makes relates to the information processing demands during performance: The information that is used during the three levels of performance can be called signals, signs, and symbols respectively. Depending on the circumstances (skill level, context in which a stimulus is perceived, or intentions and expectations of the observer), the same physical stimulus can be interpreted on different levels.

A person engaged in skill-based behavior perceives information in the form of signals: continuous sensory information about the environment that conveys no other 'meaning' than the signal itself -or its direct physical time-space data in terms of Rasmussen (1986). A stimulus that triggers a rule or activates a procedure is called a sign. Performance on the rule-based level is required to pick up signs.

Symbols refer to abstract concepts and only relate to the external world by convention. As they can only be interpreted in relation to previous knowledge, they require knowledge-based behavior.

Recently, research in the area of complex cognitive skills, converged into a theoretical framework by Van Merriënboer (1997). He developed a model for technical training. In his view, high-performance skills (HPS) are a sub category of complex cognitive skills (CCS). Although his definition does not explicitly include vehicle control tasks, and focuses predominantly on cognitive tasks[5], there is considerable overlap with Schneider's (1985) definition of high-performance skills as adapted in this thesis. A CCS includes a (large) number of constituent skills that are qualitatively different in nature, i.e.; some skills are executed in an automatic mode of processing whereas others are performed largely as controlled processes

[5] In fact, he explicitly excludes complex motor skills.

(Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Besides, there are also qualitative differences in performance between novices and experts. Van Merriënboer (1997) relates these differences to two learning processes: rule automation and schema acquisition.

These processes are connected to the distinction between recurrent skills and non-recurrent skills. Although somewhat similar to the 'closed skill - open skill' dichotomy, there is a slight difference in emphasis. Van Merriënboer relates his definition to the desired level of performance after training (i.e. skills, instead of task characteristics). In his view, even open skills should sometimes be executed as automatic processes. In fact, a particular constituent skill might be classified as non-recurrent for one training program and recurrent for another.

Recurrent skills must be performed in a rule-based fashion. By extensive training, they are subject to a process called rule automation. During this process performance becomes fast and accurate. Timesharing with other constituent skills is generally possible but skills are specific and not flexible to variations of conditions.

Non-recurrent skills on the other hand are acquired by means of schema acquisition. In the 4C/ID model[6] for technical training of complex cognitive skills by Van Merriënboer (1997) this means that organized knowledge structures must be developed that allow for elaboration and induction on information that is available to a person. Although non-recurrent skills typically are slower, and more error prone than recurrent skills they are flexibly adapted to new, unfamiliar circumstances.

In the domain of psychomotor skills, Romiszowski (1999) proposes a continuum (rather than two extremes) from reproductive to productive skills. Productive skills require the performer to produce a situation-specific response. Reproductive skills can be performed in an algorithmic manner. The value of this distinction is not situated in its innovative character (it is actually very similar to the previously mentioned dichotomies) but rather in the explicit acknowledgement of the continuum it represents (see also Norman and Shallice, 1980). This idea is namely strongly linked to the training process. Romiszowski uses the model of the skills cycle (Romiszowski, 1981; Wheatcroft, 1973) to illustrate the gradual differences between productive and reproductive skills (see *Figure 1*).

Completely automated, reproductive skills are described in terms of this model by the following sequence: 'S-1-4-R'. There is no (significant) processing of the stimulus: once perceived it directly leads to an (appropriate) action.
Skills that cannot be executed in such a reflex-like manner require recall of algorithmic procedures. This behavior calls on another step: 'S-1-2-4-R'.

For productive skills, yet another step is required. Because it is not possible at forehand to specify the desired response to a stimulus, it is necessary to plan and, if necessary, evaluate alternative plans before acting: 'S-1-2-3- (2-3-2-3-2, etc) -4-R'

This long road from stimulus to response has to be followed for 'new' skills as well. Reproductive skills can eventually be automated by means of training and instruction. If this process is completed, they are executed in the reflex-like sequence mentioned above.

[6] 4C/ID: Four-Component Instructional Design model

Romiszowski (1999) further suggests a three step instructional model to guide trainees through the different stages in the development of motor-skills: In the model, skill development starts with the acquisition of necessary knowledge to perform the task. Next, the task should be executed in a step by step manner. The last step relates to the further development of proficiency in terms of transfer of control, automatization and generalization. Transfer of control means that the action can be executed fluently, without the need to monitor ones actions visually. During automatization, the need for conscious control (thinking through the actions) is reduced. Generalization finally, is necessary to extend the skill to a continually greater range of situations.

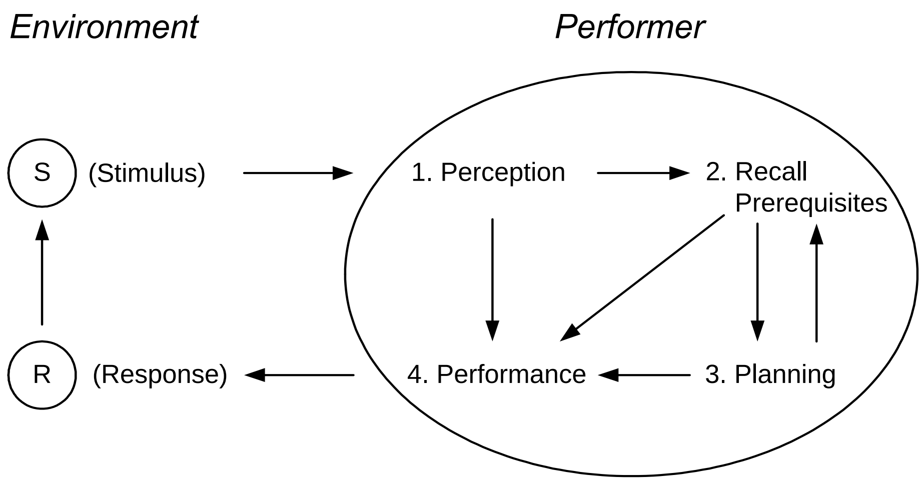


Figure 1. *Four-stage skills cycle (adapted from Romiszowski, 1999)*

Experiential methods or discovery-learning techniques are suggested for the transmission of knowledge concerning productive skills. For the acquisition of basic psychomotor skills, expository methods are recommended for both productive and reproductive skills. In the final step of the instructional process supervised and guided practice or problem solving are required for both types of skills. The concept of feedback is very important in this step. Knowledge of results and / or knowledge of performance facilitate close monitoring of the learning process especially for the fine-tuning of the reproductive aspects. Productive aspects are best supported through a process of debriefing or reflection-in-action (Romiszowski, 1999).

Stage accounts are ubiquitous in the human performance literature. They describe the learning process as a transition through two, three, or four stages. Most theories are predominantly descriptive. Apart from the starting point they provide for research, their practical usefulness for designing training or predicting training outcomes is limited. Only two models: the 4CID model (Van Merriënboer, 1997) and Romiszowski's (1999) three step instructional model in combination with the skills cycle, have been found to provide some practical handles for support of learning processes and selection of instructional techniques. Nevertheless, they do not generate useful predictions about the duration of the learning process and the mediating factors in each stage. The question is if it is possible to provide such predictions by taking a closer look at the learning process.

2.3 Learning and instruction

In this context, learning is defined as a relatively permanent change in behavior due to training[7]. The learning process cannot be perceived directly but has to be inferred from changes in behavior during and after training (usually with respect to some baseline / reference). Learning can be viewed as an optimization process where the trainee (repeatedly) tries to perform according to a set criterion. Typically, for learning to be efficient (or at all effective) trainees need specific guidance. For instance, findings from research on discovery learning (De Jong & Van Joolingen, 1998) show that trainees are often unable to systematically plan and monitor their own learning behavior.

Learning processes are affected by a large number of training and instruction factors. Training factors are factors that are related to the sequencing, frequency, spacing, duration, and content of training activities. Instruction factors are factors that are related to the support provided before, during, and after the execution of training activities. This support can take a variety of forms and can be categorized into briefing, tutoring, and debriefing activities, respectively. These supportive activities can be regarded as catalysts for the training process.

For a specific training application, the choice of a particular set of training and instruction factors defines a training program, i.e.; a controlled sequence of training activities supported by instruction. A training program can be viewed as a means to elicit / compress optimal learning experiences and as a means to control the degrees of freedom associated with learning.

In principle, the number of possibilities to design training and instruction is unlimited. The impact of training and instruction factors can be investigated from different perspectives (see also Van Merriënboer, 1997). One can look at the curriculum as a whole, the lesson, or a single event. The next sub-paragraphs provide a number of examples with regard to instructional implications of each view (see also *Table 1* from which it can be seen that decisions with regard to training and instruction have their impact on three layers: curriculum, lesson, and event.)

[7] Learning can be intentional or incidental. The latter type of learning is generally associated with experience, although this association is more one of emphasis than of principle. Many experiences are also driven by an intention to learn although this may not be the primary drive. The same applies with respect to the association between intentional learning and training: many training activities result in the acquisition of particular knowledge and skills that do not belong to the intended focus of training but rather can be regarded as incidental by-products of the training process.

Table 1: *A simplified representation of the different layers that constitute a training program.*

<table>
<tr><th colspan="9">Curriculum</th></tr>
<tr><td colspan="9">Domain, Overall training strategy, Sequencing of lessons, Criteria for progression to next lesson</td></tr>
<tr><td colspan="4">Lesson 1</td><td colspan="3">Lesson 2</td><td colspan="2">Lesson 3</td></tr>
<tr><td colspan="9">topic, general instructions, sequencing of events, duration of events</td></tr>
<tr><td>event1</td><td>event2</td><td>event1</td><td>event3</td><td>event4</td><td>event2</td><td>event3</td><td>event4</td><td>event5</td></tr>
<tr><td colspan="9">content, (de)briefing, tutoring, instructions, cues</td></tr>
</table>

2.3.1 Curriculum

For high-performance skill development often some way of part-task training is applied. The idea of part task strategies is to prevent overload, as the learner is not exposed to the full complexity of the task immediately: new parts are added one by one, as performance on the previous part(s) improves. Such strategies are a good example of an instructional manipulation at the curriculum level.

From a curriculum perspective one faces decisions that exceed single lessons. Aspects such as the structure of the training, and sequencing of lessons are emphasized. All these aspects can be considered ‘organizational’ and require specification of an overall training strategy (e.g., progressive part-task or backward chaining). Apart from that, decisions about learner progress (training until a criterion is reached or a fixed duration), and training domain and learning objectives have to be made.

There are several different ways by which a task can be divided into parts just as there are several ways to combine the tasks into training. Patrick (1992) describes these in more detail although he annotates that the empirical evidence favoring one part-task strategy over another is not overwhelming. However, with regard to the efficiency of whole-task training as opposed to part-task training (in general) the principles that have been formulated by Naylor and Briggs (1963) are widely accepted. Which training strategy will be most efficient is largely determined by two characteristics of the task: ‘task complexity’ and ‘task organization’. Task complexity refers to the demands imposed on the subject by each of the task dimensions or parts (independently). The organization of a task refers to the demands imposed due to the interrelationships between the task dimensions.

According to Naylor and Briggs, part task training and whole task training are equally effective under conditions of low task complexity. As soon as the task as a whole becomes more complex, however, for tasks of high organization (heavy demands due to interrelated task elements) whole task training becomes more efficient whereas part-task training methods yield the best results for tasks of low organization.

An example will clarify this: a task consisting of many serial steps scores low on both organization and complexity. However, as soon as some of these steps have to be executed in parallel, the organizational demands will increase although task complexity (in terms of Naylor and Briggs) does not change.

Although intuitively appealing, experimental findings sometimes contradict these principles. For example, Wightman and Sistrunk (1987) found a beneficial effect of part-task training especially for low-aptitude subjects in a simulated (aircraft) carrier-landing task. This is strange considering that carrier landing is a difficult and multidimensional task, so it would yield high scores on both organization and complexity. Findings by Gopher, Weil, and Siegel (1989) and Van Rooij (1994) in experiments using a video game (Space Fortress Game) confirm that low-ability subjects in particular benefit from part-task training. This interaction with learner characteristics may very well overrule the principles formulated earlier by Naylor and Briggs.

2.3.2 Lesson

A second perspective emphasizes the factors that affect the structure between the training activities within a single lesson (instead of between lessons). Depending on decisions taken from the curriculum perspective, the training events that make up a lesson are selected, placed in order, etc…

Patrick (1992) reviews some studies about spacing and duration of training. He states that few new insights have come up since the early sixties. The optimal duration of training seems to be dictated by practical considerations and common sense. Although in one study, massed training depressed performance during training, the effect on learning (i.e. performance after training) seemed to be small. Differences in training performance between groups tended to disappear in the retention test. Another experiment showed that the effectiveness of additional training hours diminished after certain optimal session duration. This leads Patrick (1992) to conclude that the *"…optimal combination of amount and frequency of training will vary according to the nature of the task, the trainees, the training programme etc."* (p.371).

2.3.3 Event

Zooming in on the events that make up a lesson, we arrive at the third perspective. Here, the instructor is concerned with the provision of support before, during, and after the execution of a single training activity (these are also referred to as briefing, tutoring, and debriefing respectively). Within a single training activity, the instructional process has to be adapted to the situations that occur and to the responses of the trainee. Whereas briefing and debriefing occur 'off-line' and as such can be planned in advance, tutoring can be said to be the aspect of training that is the least predictable. It comprises the reactions of the instructor concerning the specific trainee behavior ('on-line') and therefore, it is the most dynamic and time-critical aspect of instruction. Generally, workload for both trainee and instructor will be relatively high during tutoring: On the one hand, the trainee struggles to reach a higher level of proficiency. He needs the help of a tutor to focus on the right cues. The tutor on the other hand continually monitors performance and selects strategies to support the trainee optimally during this learning process.

Tutoring may include different kinds of support in different sensory modalities. Aural support includes verbal instruction, naturally occurring sounds, or symbolic ones such as beeps. Visual support can be in the form of printed text or pictures, actions of other agents, or aspects of the environment. Additionally, support can be kinesthetic or tactile in the form of a vibration or force that cues to signal a mistake, or guides the trainee towards the desired behavior.

The repertoire of a good instructor contains a multitude of tutoring strategies to support a trainee. Examples of such strategies include: correcting, confirming, explaining, intervening, prompting and providing of cues.

2.4 Conclusion

From a training point of view it is possible to look at tasks, skills and their development (whether that is in the context of high-performance or not) in many different ways. Currently there is no single approved view that guarantees success in instructional design. Each approach has positive aspects. Still, many of them are based on research on laboratory tasks or skills and do not take into account the variety of contexts a single skill could be applied in. Also the mapping of skills onto (meaningful) tasks is a difficult problem.

A considerable advantage of stage accounts is their ability to cope with the changes in skills as a result of learning. However, they lack the power to predict why, how and when such changes will take place which renders them merely descriptive. Some of them provide mechanisms that explain the transitions between stages but still are not able to predict their duration. The 4C/ID model (Van Merriënboer, 1997) and the Four-stage skills cycle (Romiszowski, 1999) provide the most practical guidelines for instructional design. Both approaches distinguish between different types of skills and provide instructional techniques that are appropriate to train each of those. The Four-stage skill cycle is particularly flexible in the way skills may change during the process of automation from productive to reproductive. This way the model is able to counter a main criticism about the rigidity of the stage accounts.

The 4C/ID model has a strong point in distinguishing between different perspectives on instruction. Depending on the level of detail an instructional designer takes (curriculum, lesson, or event), different relevant topics can be defined. The event level is concerned mostly with support before, during, and after an event, also called briefing, tutoring, and debriefing. On the lesson and curriculum levels, focus is more on the structuring / sequencing of events or lessons respectively.

This levels view is adopted here even though the definition of high-performance tasks in the 4C/ID model is different from the one used in this thesis. Such differences should be set apart because there is a strong need for a comprehensive theoretical framework that helps to predict the duration of the learning process as a function of skill level, task characteristics, interventions, and other mediating factors. As the current models do not seem to be able to provide the answer because of the complexity of the learning process, this thesis will zoom in on the learning process at the lowest (event) level.

Just as each perspective or level focuses on different instructional aspects, it can be asserted that each perspective has consequences for the medium that is used to convey training and instruction. The characteristics of simulators are especially appropriate to compare different forms of tutoring. Therefore, this topic will be discussed in the next chapter.

Chapter 3

Simulators

Abstract
Once a training simulator is properly equipped with the tools for performance measurement and feedback, and specific instructor support facilities such as an instructor console, it has a number of didactical advantages over training in an operational environment (e.g., a real car). The training simulator is especially flexible when it comes to control over content, structure and timing of instruction. Furthermore it offers possibilities to standardize training program content while differentiating the amount of instruction or practice for individual trainees. One aspect that receives special attention in this chapter is augmentation of reality: the provision of artificial cues to help the trainee focus on the relevant information. This implies that for training to be efficient it is not always desired to maximize fidelity.

3.1 Introduction

When simulators are divided into different categories (e.g., Patrick, 1992; Stanton, 1996) two main categories are generally distinguished: research simulators, and training simulators. Research simulators are used for example in the design of buildings, vehicles, etc. or in the evaluation of procedures. The data from the simulation process are used for the analysis of the forces that act on a structure, aerodynamic properties of an object, or the efficiency of performance. Training simulators are deliberately designed for the provision of training. They should incorporate special features to support the instructional process, and to evaluate human performance. Trainees interact with the simulator through predefined scenarios of increasing difficulty while their performance is monitored and feedback is provided. An instructor (system) is responsible for monitoring trainee progress and selecting exercises to support optimal learning.

Experimental research investigating the effectiveness of instruction in a (training) simulator is not easily found. A great deal of training simulator studies (especially older ones) has predominantly been concerned with technical aspects of these devices. As it can be foreseen that technical and hardware aspects will cease to be the major cost drivers for simulators, it is expected that instructional design and development of software for training will become increasingly important to make simulation cost- and training effective. However, so far, studies that investigated instruction in a simulator are often very narrow i.e., completely focused at development of a training program for a single application or a specific task (e.g., Wolz, McKeown, & Kaiser, 1990; Ramesh & Sylla, 1990; Gordon, Babbitt, Bell, & Sorensen 1994). Because the results of these studies are so specific, generalizations can only be drawn from them to a limited extent. Therefore it remains difficult to distill principles for the scientific design of simulator instruction and training programs from this line of research.

In principle, any training simulator can be said to consist of the same basic components (see *Figure 2*). How exactly these components will look depends largely on the nature of the training task. Compared to (PC-based) procedural simulations, the dynamic nature of vehicle control poses additional requirements to the simulator equipment in structural terms (control interface, and accessory systems). Nevertheless, it must be said that the difference between PC-based training simulation and a training simulator is a matter of emphasis. The term simulator is associated more with hardware components whereas simulation emphasizes a process. In this sense, both concepts (simulation and simulator) will be used next to each other.

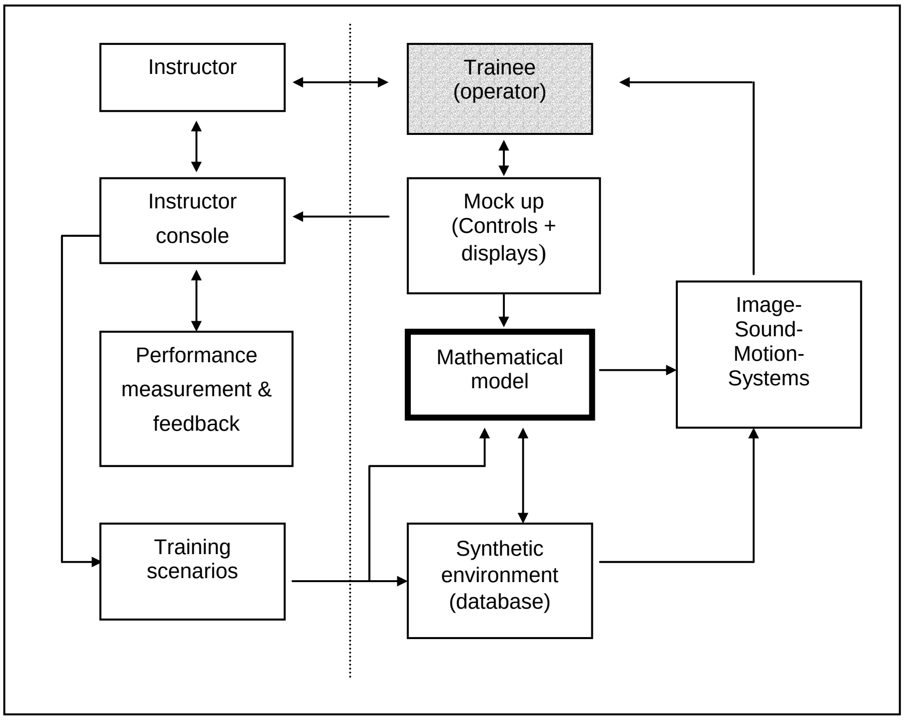


Figure 2*: Generic simulator configuration. The components that are placed to the left of the dotted line are particularly specific to a training simulator (as compared to a research simulator). The arrows indicate the interactions between the different elements*

At the heart of any simulation lies the mathematical model: a collection of equations that specify the behavior of the simulated system in reaction to input generated by the operator or by the changing conditions in the environment.

The operator perceives the (virtual) environment through the image- sound- and motion systems. These ‘sensory systems’ serve as the windows on the synthetic environment that is represented in a database. Not all of the sensory systems have to be present but at least some of their information is required to perform the task. In this case, more information generally means higher perceived realism.

The trainee generates his (her) input in a mock-up of the system: a representation of the controls and displays of the operational system. This can be anything from an exact copy to a software representation on a computer screen. For motor skill learning the relevant controls and their control force loading ('touch') should be represented accurately whereas a relatively simple software / keyboard representation may suffice for many procedural training tasks.

Whenever the input to the controls is processed through the mathematical model, the mock-up as well as the sensory systems of the simulator should react correspondingly thus creating an interaction between trainee and simulator.

So far, it may seem as if a simulator is nothing more than a representation (dynamic model) of an operational system. To 'change' it into a training device, a number of instructional aids (including tools for: scenario generation, performance measurement, and delivery of feedback) are required. These are usually referred to as 'Instructor Support Systems' (ISSs). An ISS generally includes an instructor console from which information from control elements (mock-up) about performance (acting with the system) can be read. A more advanced ISS will also include a performance marking and feedback (PMF) system. The PMF facilitates automated registration and interpretation of performance and provides its input to the instructor console (or a printer).

Via the instructor console of the ISS the selection of training scenarios can be done. External events are generated using a database of this environment. Events are presented to the trainee / operator in the form of pre-specified scenarios. Alternatively, (semi-) intelligent agents may operate autonomously in the environment.

Simulator training will not be optimally efficient as long as such instructional facilities are not present or if they are designed poorly. This will reduce beneficiary contributions that could result from simulator specific aspects of the training. In fact it leads to an approach of training that is essentially the same as training on the operational system. A replication of the standard training, however, implies that the weaknesses of the original training program are reproduced together with its favorable aspects. The use of simulation, on the other hand, offers the possibility to improve on the sub-optimal training aspects of the operational system to end up with the best of both worlds. Taking into account that any aspect of a simulation can be changed, effort should be invested to optimize sub-optimal training aspects in particular.

However, this notion is yet to gain common acceptance. Only recently, Verstegen, Barnard, and Van Rooij (1999) collected information on a total number of 39 military simulator facilities throughout Europe. In their analysis of the current use of training simulators, it was concluded that the possibilities of simulators were not used to their full extent. In particular facilities for instructor support, provision of feedback, and registration and assessment of performance were found to be either poor or lacking in most of them.

In another study, the effect of simulator training for Leopard II tank drivers was found to be detrimental to performance on the operational system, that is: negative transfer occurred (Van Breda & Boer, 1988). Although this was attributed mainly to the physical aspects of the simulator (visual-, and motion systems) it was also concluded that the instructor facilities, and the training program and scenario's were poorly designed. Korteling (1990; 1991) showed that the instructor console lacked some of the necessary facilities for provision of feedback and measurement of performance. Furthermore, it had a poor ergonomic design, and computer generated output with regard to trainee performance was not easily interpretable. After a drastic revision of the simulator including changes to the motion- and visual systems, changes in the training program, and redesign of the instructor console and performance measurement and feedback system the simulator yielded positive transfer (Veltman & Korteling, 1993).

3.2 Advantages of simulator training

In the literature, different arguments are used to point out the value of simulators for training high-performance tasks. Three of these appear (in some form or another) in the literature on simulation repeatedly (e.g., Hays & Singer, 1989; Patrick, 1992; Stanton, 1996). These pertain to costs, availability, and safety.

- Often the operational system is too expensive to be used for training due to the increased wear and tear that is incurred during training or due to fuel expenditure. Especially for heavy (military) equipment such as aircraft and tanks, the use of simulators has created training possibilities where training would not have been possible otherwise.

- When the real system is not available because of maintenance duties or operational deployment training must be arranged otherwise. This argument applies to trains or other means of public transportation by rail. In the busy schedules, often no time is available for training of drivers on the available tracks. A related aspect pertains to the environmental conditions necessary for training. These might not occur frequently enough, might not be available, or are difficult to realize. Adverse weather conditions such as snow, heavy rain, or fog, and special situations such as bad road conditions are often not encountered in the course of a person’s driver training. Similarly, astronaut training can only be done on earth.

- Finally, safety- and / or environmental regulations may preclude the training of particular tasks. Even if the circumstances would be available sometimes it is simply too dangerous to train a certain situation in reality, e.g., engine failure on an oil tanker in heavy weather or a truck that 'jack-knifes'.

With these three practical advantages in mind, it can be seen that the use of simulators can lead to increased training-efficiency and offers enhanced training possibilities compared to (conventional) training on the real system.

Meanwhile, it is often overlooked that training in a simulated environment also offers a number of didactic advantages. These are concerned with the content and form of the *simulator training* (as opposed to the *training simulator*) and include the following aspects: control, standardization, differentiation, performance registration, and augmentation.

Although some of these aspects can also be manipulated in operational settings when high-performance tasks are considered, a simulator is far more flexible.

3.2.1 Control

Simulation of a process gives one control over content, structure, and timing of the instructional process at all levels of instruction. From a 'lesson perspective' (see

Chapter 2), control of content can be used to provide the trainee with an environment that is challenging though not overwhelming when the real system would be too complex to handle during the initial stages of training. Likewise, it is possible to increase difficulty of the exercises gradually in the progress of training (curriculum perspective). Although this is possible during training on the real system to a certain extent, the flexibility of a simulator is far greater and in reality, unpredictable and uncontrollable factors can always disturb a carefully planned training scenario.

In addition, complex tasks can be broken into smaller parts that can be practiced in isolation during initial training (part-task training). There are many different ways a task can be broken into parts. However, it goes beyond the scope of this thesis to discuss them in more detail. Extensive reviews can be found for example in Patrick (1992) or Farmer et al. (1999).

Finally, concerning timing, the delivery of training events can be arranged in such a way that maximum use is made of (limited) training time. This way, pilots who specifically want to train their landing skills do not lose time by preparing their aircraft, and getting up in the air each time. The increased number of approaches that can be practiced in the simulator presents an opportunity to spend training time much more efficient.

3.2.2 Standardization

Control of content is closely related to standardization. On a simulator, training (content) can be standardized more easily than on the operational system. This way, the time of day, season, or weather conditions do not affect (limit) the training conditions the trainee is confronted with. Any trainee will be able to practice the full gamut of training activities. When a sufficiently large variety of situations (or events) are present in a simulator-training program, it can be assured that every trainee has received instruction on all relevant topics by the end of the training.

3.2.3 Differentiation

Within a training session, differentiation can be desired, as some trainees need more instruction than others do. Trainees with a need for extra practice trials can repeat either the same or a similar exercise until criterion performance is reached. Replay of recorded scenarios provides the opportunity to analyze trainee performance in detail and give extended feedback. Although it may seem that the idea of differentiation conflicts with the concept of standardization, at closer look this is not the case. Differentiation is a principle that is applied within a lesson and results in ‘more (or less) of the same’ depending on the needs of the trainee. Standardization is concerned with the learning content over the curriculum. In other words all trainees receive instruction on the same topics (standardization) but some need more practice than others do (differentiation).

3.2.4 Performance registration

An important didactical advantage of simulators is the possibility to store and analyze all kinds of performance data automatically. This enables both grading and continuous performance monitoring in a standardized objective way.

Based on such objective data, instructors will be able to provide accurate feedback with relatively little effort. The direct availability of objective performance data will also facilitate the (partial) automation of feedback generation by the simulator.

Obviously, these options will affect the role of the instructor. By (partly) automating the generation of feedback, it will shift from an active to a more passive, monitoring, supervisory, and evaluative role. This will either lead to a significant reduction of workload for the instructor or an increase in the number of trainees an instructor can assist simultaneously (provided that the facilities for the instructor; the instructor operator station (IOS) support this).

3.2.5 Augmentation

While the above advantages mainly pertained to the curriculum or lesson perspective, it is also possible to show benefits from simulators at the event perspective, for example by means of augmentation of reality or augmented cueing as it is also called.

The term augmentation refers to the presentation of artificial information (that would not be present otherwise) to a user of a particular system or device. The information from an augmented cue should be easy to interpret and ‘automatically’ lead someone to take the right actions. An environment that is manipulated using these kinds of techniques for instructional goals is called an augmented reality (Young, Stedmon, & Cook, 1999).

Using a simulator it is relatively easy to manipulate (virtual) reality by providing augmented- or artificial cues and feedback. In principle, there are innumerable ways to adapt a task environment by means of augmented cues and feedback. Augmentation can be presented aurally in the form of beeps, alerts, or noises associated with system failure. Visual augmentation can be realized by means of lights, color changes, arrows, or objects. For example, the markings on a virtual road can be programmed to light up as soon as an automobilist crosses them and is in danger of getting of the road. Additionally, a sound could be generated to warn the driver. This is in fact what happens on some (real) roads where the normal rustling of the tires changes into a high resonant tone when they are brought into contact with the road markings (that are specially textured for this means). Another form of augmentation could be presented by (vibro-) tactile or kinesthetic stimuli. This category of cues is directed at the sense of touch. An example would be a short computer generated jerk or vibration at the steering wheel signaling the appropriate moment for steering into a curve.

Augmentation of reality can serve two purposes: in an operational environment it can help employees (e.g. pilots, or ATC operators) to improve their performance or lighten their duties; during training it can be used as an instructional aid. In the latter case augmentation is very similar to tutoring albeit that the cues are not provided verbally but are presented in a virtual environment. Just as with other forms of instruction it is important not to provide the augmented cues continuously. This might result in a situation where trainees become dependent on the augmented cues and will be helpless as soon as the cues are withdrawn[8].

The presentation of an optimal flight path to an aircraft pilot is an example of augmentation as a support tool under (normal) operational conditions. Such a 'highway in the sky' helps aircraft pilots during descent by depicting a flight path on a display. As shown in Figure 3, the display provides information about the best way to approach the runway. Strictly speaking, a speedometer in a car is also a form of augmentation: the abstract concept of ‘kilometers per hour’ cannot be determined from the environment very accurately. However, it is not used as a training aid because the information remains available even after a person is licensed to drive. Instructional augmentation on the other hand is meant to emphasize relevant information only during training.

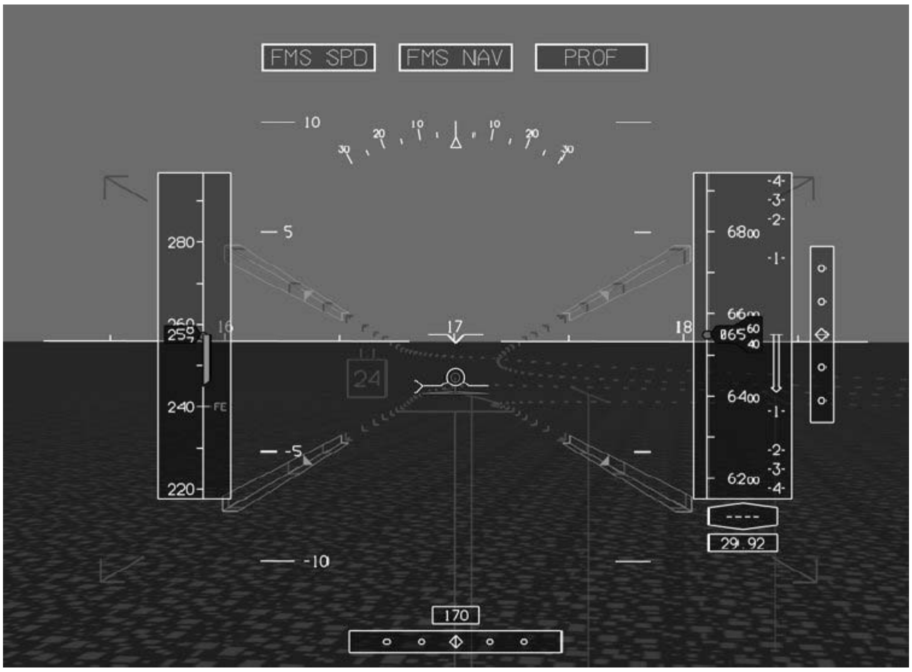


Figure 3. *Augmented cues in use in an aircraft display. The so called 'highway in the sky' helps aircraft pilots during descent by depicting a flight path on a display.*

Source: http://techreports.larc.nasa.gov/ltrs/PDF/2003/cp/NASA-2003-cp212164.pdf.

[8] By gradually decreasing the amount of feedback as performance improves during training (a process called 'fading') this problem is avoided.

It has been suggested that augmented cueing could be especially useful for training novices although some studies also find positive results for more experienced trainees. As compared to expert performers, novices are not very efficient in sampling the relevant information from the environment (Ericsson & Lehmann, 1996). This is one of the causes for the differences in performance between them. By pointing to (and emphasizing) the relevant cues novices are expected to learn what is a relevant information and what is not so that after some training, the learner should be able to focus on these cues himself.

In a program aimed at identification of optimum design and instructional features for flight training simulators Lintern, Thomley-Yates, Nelson, and Roscoe, (1987) found that beginning flight students as well as more experienced pilots benefited from simulator training with display augmentation albeit in a different way. Subjects had to fly over a simulated landscape and drop a bomb on a target that was designated somewhere in the scene. For half of the subjects target acquisition was aided by visual guidance (off target augmented feedback): if one of these pilots exceeded certain error limits based upon an optimal flight path, yellow cubes that were located on the desired flight path were displayed. After a training phase a number of transfer trials were flown without augmented feedback. It turned out that only the less experienced pilots benefited from the augmented feedback with regard to pitch control [9] both in training and during transfer. With regard to bombing accuracy however, the more experienced pilots showed an increase in performance compared to a decrease for the less experienced subjects. Lintern and his colleagues attributed these results to the difference in sensitivity of performance measures to different stages of learning. Since the cues were primarily designed to improve diving parameters and only secondly to improve bomb releasing accuracy the less experienced pilots only improved on the diving part of the task. On the other hand, the more experienced pilots could not benefit much from the information considered that they already mastered this aspect of the task. Therefore, they were able to focus on (and improve their skills with regard to) the accurate delivery of the bomb.

Subsequent experiments (Lintern, Roscoe, Koonce, & Segal, 1990; Lintern & Koonce, 1992) gave further support for the value of visual augmentation for transfer of contact flight skills. They showed that visual augmentation can enhance the acquisition of visually supported skills and the findings add weight to the previously discussed experiment. Also, the data from the latter study indicated that subjects were able to benefit most from augmented cues in an artificially degraded environment. In other words, augmented cues seemed to counteract the disadvantages of training in an environment with low visual scene detail[10]. Since the use of visual augmentation is considerably less expensive than providing high (or moderate) scene detail, this could be seen as convincing support for low-cost simulation approaches. However, caution is needed in the interpretation of these results because degraded environments may also lack the cues that are necessary for normal performance as a result of which subjects learn different skills than desired.

Summarizing, it seems that some forms of augmentation work in some settings. The way in which this happens is complex because training effects do not always show

[9] Elevation of the nose of the aircraft
[10] Transfer was measured in a moderate-detail flight simulator

up during transfer. In line with this remark, Schmidt and Bjork (1992) state that better performance during training does not always mean better retention of skills. They indicate that still a large number of questions remain to be answered in the area of feedback and transfer.

3.3 Fidelity and validity

Unfortunately, there are limitations to what can be simulated. These limitations may reduce the perceived realism (face validity) of a training simulator and, hence, its acceptance as a training device by trainees and instructors. Also these limitations may reduce or impede the transfer of skills from the training simulator to the real system if critical cues for performance cannot be simulated correctly.

In addition, the cost of training simulators may be very substantial. Procurement costs of training simulators often exceed the procurement costs of the real system; the same applies to the costs of operating and maintaining training simulators. For this reason, much effort is being invested in investigating ways to optimize the benefit / cost ratio of training simulators (Orlansky, 1989; Boldovici, 1987). However, it should be noted that empirical benefit / cost studies are relatively rare. Most of the studies that are reported in the literature focus on issues of fidelity, i.e., the extent to which the behavior of the simulator mimics the behavior of the operational system.

The comparison and assessment of the results of these studies is complicated by several factors, the lack of consistency in the definition of fidelity being the most prominent. In particular, older studies and studies in the field of engineering (e.g., Allen, Hays, & Buffardi, 1986; Thomson, 1989) apparently fail to appreciate the fact that it is not the fidelity of the simulator that is the goal of simulation but the efficient transfer of training. Although these goals are certainly related, they are by no means the same (Lintern, Sheppard, Parker, Yates, & Nolan, 1989; Roscoe, 1991). After reviewing the available literature, Korteling, Van den Bosch, and Van Emmerik (1997) propose the following subdivision and definitions of the term fidelity:

Physical fidelity pertains to the similarity between the operational system and the simulator. A lot of factors contribute to the physical fidelity: mock-up, displays and controls, sound and vision, motion, accuracy of the mathematical model. Usually this includes the appearance of the simulator or rather, the mock-up. (This latter aspect is also called face validity).

Psychological or *functional fidelity* is the degree of similarity between trainee behavior on the simulator and on the real system. In this view, it does not matter if the simulator is made out of cardboard[11] or if it is a complete copy of the operational system, as long as it elicits the desired behavior. In most cases then, it is clearly sub-optimal to strive for the highest possible level of physical fidelity in simulation from a training point of view (Lintern et al., 1989; Patrick, 1992). The level of physical fidelity needed to achieve functional fidelity depends on a variety

---

[11] In an experiment by Prophet and Boyd (1970) cited in Patrick (1992) an aircraft model made of plywood and photographs was as effective as the real aircraft for training start-up and shut-down procedures.

of factors. Boer (1991) mentions the type of task to be trained, the proficiency level of the trainees, the difference between criterion performance and maximum performance. It might be expected that didactic factors will also play a part in this relation. No matter how sophisticated a simulator is designed, without an appropriate training program, it cannot be used efficiently.

This discussion about fidelity should lead to one observation: The important part of training is that people learn what they are supposed to learn. Preferably as efficient as possible. This is in short the definition of validity. However easily defined, it does not simplify the debate.

Validity is a complex concept. It is affected by functional fidelity, quality of training, the type of task being trained, and trainee level (Korteling et al., 1997). Talking about 'the validity of a certain simulator' then creates a certain sense of discomfort, for the validity of the simulator will change as the trainees gain experience or as other trainees will use the device. Taken to the extreme, a simulator will have multiple 'validities' each referring to a different combination of the factors that affect the concept.

Is concept of validity futile, then? No. From a slightly more pragmatic point of view, validity *can* be determined, for example, by measuring transfer of training. In carefully designed experiments it is possible to measure the extent to which learning a certain training task influences performance on an operational task. Still this is not a simple endeavor. Korteling and Sluimer (1999) provide an extensive overview of the pitfalls (artifacts) that may affect the result of transfer studies. One of their conclusions is that *"Speaking about the validity of a simulator only makes sense if functional aspects are taken into account, such as the purpose of the simulator, the tasks, trainees (or subjects), training methods, and additional training aids. […] For training simulator studies, the potential benefits of additional instructional facilities […] must be taken into consideration [during the process of validation]."*

Notwithstanding the methodological problems associated with transfer of training studies, once people realize that there is more to reaching training objectives than fidelity alone, a large step has been made to better use of simulators in training.

3.4 Discussion and Conclusion

Based on the above, it can be said that one of the reasons to use simulators for training high-performance tasks is that they can contribute to a significant improvement of training efficiency. However, existing simulators and their advanced instructional features are not used to their full potential (Polzella, 1983; Verstegen, Barnard, & Van Rooij, 1999). The reason for this is the failure to appreciate a simulator as an *advanced training tool*. Instead, many people expect simulators to operate exactly -and only- as the real system. Two points in particular should be made against this fallacious assumption:

First, training does not always require the same conditions as operational performance (Schneider & Detweiler, 1986). Simply because not all aspects of operational performance have to be trained to the same extent. Just as a marathon athlete does not train by running a marathon each day, correspondingly, a simulator may be a valuable training device by offering ways to practice a single aspect of performance. Following this approach, it will be possible to practice some aspects ten or twenty times in a row without loss of time to find a similar situation again (e.g. when training to 'pull out of a highway'). In a real car, you have to drive back repeatedly to access the highway and drive to the next exit where you (finally) can execute the desired maneuver. From an instructional point of view, this is a waste of time. Compared to reality a simulator can provide: better sequencing of instructional events (and thus prevent workload problems), more and better feedback with help of data storage and scenario re-runs, and more frequent exposure to educational situations.

Second, to provide efficient training on a simulator it may be better to train only a few parts of the task in the virtual environment and other parts on the real system. This pragmatic point of view corresponds with the ELSTAR[12] approach to simulation (Korteling, Helsdingen, & Von Baeyer, 2000). Some tasks cannot be trained efficiently with a simulator. For some tasks the costs of simulating would exceed the benefits (e.g., terrain driving with a cross-country vehicle requires 3-D vision, accurate terrain representation along three axes, accurate motion cueing with at least three degrees of freedom, etc. because this is essential for judging terrain characteristics. The costs of including all these aspects in a simulator -if possible at all- are very high).

ELSTAR propagates that the concept of full-mission simulators is simply not feasible for many high-performance tasks. According to the ELSTAR philosophy, one should use a simulator to train only those tasks (or sub tasks) that can be trained with maximum effectiveness and minimum costs. All other tasks should be trained on the real system. This requires a different approach to training development (see also Chapter 4.3 on Low cost driving simulation).

In this view, training efficiency is not a (simple) derivative of the physical fidelity of the simulator. A simulator or part task simulator can only be efficient when it provides a relevant contribution to (some aspects of) a training program.

A word of caution has to be spoken about part-task, and low-cost simulators. If a training device deviates from reality too much or if the training situation bears too little resemblance to real training circumstances, it may be difficult to motivate trainees or instructors to participate in training. For example, in their validation study of the TNO-HF[13] low cost simulator Sluimer and Van den Bosch (2001) found indications that the instructors lacked confidence in the simulator as a training device. This opinion is likely to have biased their judgments of trainee performance. This calls for a certain level of 'face validity' in a simulator even if this is not deemed necessary from a purely theoretical 'educational-technology' point of view.

Much research is necessary before these training devices will be able to live up to the expectations. Such research should be focused on the use of the advanced instructional features of these training devices considering that the main goal of an instructional tool is not the representation of reality but the provision of optimal learning conditions.

[12] European Low cost Simulation Technology for the ARmed forces.
[13] TNO-HF The Netherlands Organisation for Applied Scientific Research - Human Factors

Roscoe (1991) has already stressed that the optimal conditions for learning are not necessarily equal to the most favorable conditions for ‘normal’ operation. Straightforward as it may seem, in the latter case the goal is not to optimize learning but to optimize system performance. Learners typically benefit from the errors they make if proper feedback is provided. In operational settings however, generally no facilities for this kind of feedback are present because operators are not expected to make errors. Apart from that, the real world does not place instructional situations under the control of an instructor. In this light, the approach to simulation and training that is to be followed requires training designers to deliberately deviate from reality. Some researchers even take this idea one step further and encourage us to make a training environment ‘as phony as can be’ (Roscoe, 1991) on the condition that the optimal circumstances for learning are created.

Deviation from reality can be conceived for each perspective on training (curriculum, lesson, or event). For example, it is possible to adapt and control the order in which the tasks are learned (sequencing of tasks and lessons). One could for example think of practicing gear shifting at low speeds without any other traffic initially just to make an insecure trainee comfortable with these actions. Speed and traffic intensity can be increased gradually. Such an approach, which is similar to part-task training strategies would be difficult (if possible at all) to realize in practice.

With regard to lessons, scenarios can deviate from practice in the number of learning moments that are presented. Events can be reproduced and repeated, long distances / time intervals between interesting situations can be eliminated to increase the number of instructional events (sequencing of events within a scenario).

The tutoring aspects (within an event) can also be adapted. For example by changing the ways (modalities) in which tutoring occurs. Also, deviation from reality can be helpful by providing a focus (emphasize) on relevant cues (i.e. highlighting certain aspects of the environment), or creating artificial cues (augmented cueing and feedback). This principle of augmentation is explored throughout the experimental part of this thesis because the concept has not received systematic attention in spite of its intuitive appeal.

Chapter 4

Driving in a Simulator

Abstract

It is argued that car driving is essentially an everyday life example of a high-performance skill. Even though formal instruction is relatively short, drivers continue to improve on their skills over a long period of time. A (driving) simulator could therefore be a valuable tool to improve the efficiency of formal driver training. Yet, commercial driving simulators have not been around in training for more than a few years - most likely because of their unfavorable cost-efficiency ratios. By cleverly allocating the training tasks between simulator and real system (car), however, the efficiency of training can be improved and simulator costs can be reduced. This is called low-cost simulation. A simulator based on this approach was used as experimental environment for the current research. Its structure and components are described to conclude this chapter.

4.1 Introduction

The empirical part of this thesis is about driver training in a simulator[14]. Car driving involves perceptual motor skills, procedural skills, and cognitive skills that have to be time-shared. With the increasing density of both traffic and information to attend to, the dynamics of the environment impose a heavy load on the driver.

In the Netherlands, the instructional process is taken care of by specialized instructors in specialized schools. Formal training takes on average 35 hours (Kuiken, 1996). Although only one third of the students passes their exams at their first attempt, most people get their license eventually. However, the learning process continues for a long period after that (Duncan, Williams, & Brown, 1991). Kuiken (1996) quotes European research (OECD, 1976) to state that drivers continue to improve on traffic insight over a period of years and a traveled distance of over 100.000 kilometers. In line with this finding it is a well-known fact that inexperienced drivers are responsible for a comparatively large part of the total number of traffic accidents (e.g., Regan, Deery, & Triggs, 1998). This is partly caused by the fact that most of the circumstances leading to an accident never will have occurred during driver training -because of the short duration of training and evidently because the instructor will intervene before a situation turns into a real emergency-[15].

From the facts above it becomes clear that car driving is essentially an everyday life example of a high-performance skill. This implies that the use of a driving simulator for training presents a valuable contribution, both to speed up the development of driving skills as well as to provide a safe environment to learn from ones mistakes (Ivancic & Hesketh, 2000; Wachtel, 1996). Despite these high expectations, it was not until 2001 before the first driving simulators were (commercially) used in driver training (Kappé, Van Winsum, & Van Wolffelaar, 2002).

4.2 Driving simulation

The most impressive examples of driving simulators have been designed for research and development purposes. An example of such a powerful simulator is the National Advanced Driving Simulator (NADS) in Iowa (http://www.nads-sc.uiowa.edu/). Theoretically, all aspects of car driving could be trained in these devices. However, they are not designed for training and because of their high costs, commercial exploitation cannot be cost-effective by far.

Despite this observation, the number of cost-effective driving simulators for training slowly increases. For example, in a paper by Thoeni (1999) an experiment is described to train truck drivers for the Swiss Army. In this experiment, a group of drivers was trained on a simulator for half of their training time. Sessions

[14] Unless explicitly stated otherwise, in this thesis, the term driving is used in the context of passenger cars.
[15] Another factor is that young (male) drivers are known to overestimate their driving skills and do not recognize risky situations as soon as more experienced automobile drivers do (Decina et al. 1996).

in the simulator were alternated with 'conventional' training sessions (i.e., on the road with a driving instructor). These made up the other half of the training. After six weeks, the performance of these drivers was almost identical to that of a control group, which had received training in the conventional way only. From this it was concluded that the application of simulators (even for a part of training) could lead to a substantial reduction of training cost compared to conventional driver training. Considering that one instructor may be able to supervise up to five simulators simultaneously, further increases in training efficiency for the future can be expected.

Regardless of its positive findings, unfortunately this study does not provide any information about the training program in the simulator as compared to the training on the real truck. It is therefore assumed that the simulator training was kept as similar as possible to real truck training.

This seems to be symptomatic of the experiments that are reported in the literature: although most studies report the characteristics of the simulator that was used, only few experiments provide a detailed description of what the training program constitutes. If not necessary, it would be at least interesting to know how the simulator was used, and what specific didactic features have (and have not) been used in the training. This would help us to increase the efficiency of simulator based training programs.

Wierda (1993) conducted such an experiment in which he explored the use of simulator-specific ways of providing instruction (and clearly reported how). In this study, six subjects without prior driving experience received instruction in a driving simulator. Apart from more traditional instructional methods, three simulator-specific training features were used: The helicopter view -as it was called-showed the context of the car in a tethered view from behind the own car. This way it was attempted to enhance insight in traffic situations and review ones behavior as an objective observer. The training program could be stopped (freeze mode) when provision of additional instructions was needed, and furthermore it was possible to replay the last 30 seconds of a scenario (from the perspective of the own car).

The reactions from the trainees as well as those from the examiner with regard to the simulator lessons were positive. After six lessons of 1.5 hour in which 13 learning goals were trained, an examiner (expert driving instructor) observed the subjects driving in the simulator. According to him, five trainees would have had the experience of at least 20 ‘normal’ driving lessons. One subject was rated as a 'significantly less experienced driver'. Additionally, two students who did not follow the complete course were also judged as 'less experienced'. This makes it plausible that examiner judgment had at least some relative validity. However, and this is a weak point of the study, only one person judged the trainees and they were only observed while driving in the simulator. Furthermore, only six subjects completed the full course. Another important objection is the fact that the examiner knew that the subjects had been training in a simulator and not on the road. Furthermore, no statements about the validity of the simulator with regard to the driving tasks were made.

Although there were some serious methodological problems with this study, there are two important positive aspects to it: subjects learned a lot and enjoyed it; and, most important, by exploiting the didactic possibilities the simulator offered, training became more efficient[16] than standard driver training.

4.3 Driver training in a simulator

Despite the positive developments in the field, complete driver training does not yet occur in a simulator. The scale on which simulators are used for part task driver training is small, especially when compared to the use of flight simulators. The obvious question is why. For aircraft training at all levels, simulators have become indispensable to the extent that it is now inconceivable that someone becomes a civil or military pilot without ever having experienced a simulator during some phase of training. In fact, when it concerns conversion training, it has been possible for over ten years to get licensed without training in the real aircraft - this is what Powell (1990) calls 'zero aircraft flight time'.

The answer probably is that compared to flying, the driving task has a number of characteristics that present a challenge to the simulator. For example, the amount of moving objects in close range that needs to be dealt with is considerably higher in driving than in flying. Because of the high angular velocities of these objects, the update frequency of a driving simulator needs to be high to ensure smooth transitions of one image to another. The more objects that are present in the immediate environment, the more computational power is required.

The need for a detailed scene sets high demands for visual database developers too. Compared to a commercial aircraft pilot who usually cruises at 30.000 ft. car drivers will get a close look at other cars, the road, houses, and trees. All these elements in the environment have to be modeled, scaled, and positioned more or less correctly to give the driver some sense of reality. Apart from that, the behavior of all traffic participants must be valid for the simulation to be convincing. Whereas the 'occasional' aircraft that is encountered in a flight simulation could be controlled by a human player, the number of cars, pedestrians, cyclists in a realistic road environment is too high to have each of them be controlled by an instructor or role player. To ensure that all these entities perform human-like (they might even break a rule now and then), you have to rely on 'agent technology'. Agents are bits of software that are able to act autonomously, they respond to events and processes in their environment and may interact with other agents. Once an agent is able to perform these actions in a convincing (believable) way, its behavior will appear to be intelligent (Masthoff, 1997).

One way to deal with the challenge of creating a convincing environment is to use videodisc or DVD images in a simulator. The main disadvantage of this type of simulation is the limited flexibility and the lack of interactivity of the driver with the environment. The use of this kind of simulation is probably limited to activities such as risk-awareness training (e.g., to show a person what might go wrong when driving with high speed through a built up area). DVD-based simulation is almost certain to generate a major shock effect when a child suddenly crosses the street. However, the possibilities to control the car are rather narrow (mainly limited to

[16] as rated by an experienced driving instructor.

acceleration and braking). For that matter, computer generated images (CGI) are much easier to manipulate and the computer power they require has come within reach of desktop machines recently. In the world of gaming, it is clearly visible that each new generation of computers entails increases in power and speed by several orders of magnitude compared to the previous generation. Even low-cost personal computers are able to display stunning 3D graphics and the virtual adversaries (agents) that we encounter become increasingly intelligent. Although the behavior that is appropriate in most of these games is not likely to be encouraged in motorists, still these games can give an impression of the role personal computers might play in commercial driver training.

Several companies have acknowledged this and have developed software for driver training at home (e.g., Sierra, 1999). The potential of such software for driver training, however, must be considered limited. A desktop PC, even if it is equipped with a steering wheel, pedals, and a gear stick, is not suitable for full driver training because the display offers only a restricted field of view (FOV). At an approximate viewing distance of 55 cm a standard 17'-monitor covers only about 35° of the horizontal visual field. This is insufficient for a large number of driving tasks such as looking into crossroads for other traffic, overtaking other cars, and orientation (Kappé & Korteling, 1998). Therefore, their usefulness is limited to basic procedure- and theory training. Since field of view is such an important factor in learning to drive, a driving simulator should be equipped with facilities to display at least 180° field of view and side, and rear-view mirrors to be useful for training traffic participation and traffic insight. For training of fine maneuvering (parking, turning on the road, driving backwards, etc…) the preferred FOV will even approach 360°. Image generation therefore, will remain one of the major (hardware) cost drivers for effective driving simulation of these tasks.

For these latter tasks, motion cueing is another cost driver. The subtle motion information that the driver receives from the road curvature, gusts of wind, acceleration or deceleration is hard to mimic in a simulator. This requires a motion platform with (at least) 6 degrees of freedom. Apart from the costs for a so-called 'hexapod' (the six legged hydraulic or electric apparatus to simulate motions of the cabin), this incurs extra costs e.g. for a solid concrete foundation to cope with all the forces that are generated during operation. Finally, the process of tuning a motion model is a difficult one that is more important than it may seem at first glance: if the motion cues are not in accordance with the visual stimuli in the simulator, motion sickness is likely to decrease the performance of trainees.

4.4 Low cost driving simulation

These plain observations might lead to the conclusion that cost-effective driver training in a simulator simply cannot be realized. However, some researchers have adopted a different approach. Boldovici (1992) for example dares to doubt the benefit of moving bases for most applications. He contends that people often do not notice it when motion cues are switched off during training. Apparently, for many tasks motion is not a critical cue so it does not have to be simulated. Similarly, there are many tasks that do not require a 360° field of view (Van Winsum & Korteling, 1998). Using this kind of knowledge to determine which training tasks should be simulated or not can reduce the training cost with an order of magnitude. This is what is called 'low-cost simulation' (Korteling, Van den Bosch, & Van Emmerik, 1997).

The starting point for such an approach is a thorough task-analysis of the domain. For each sub-task that is distinguished, a number of critical cues have to be defined. The nature of these cues is used to determine whether a sub task can be simulated with relatively little effort (and costs), or not. If we state for example that motion cues are too expensive to incorporate in a low cost simulator, it has to be decided to train those tasks for which motion is a critical cue on the real system (car) instead of on the simulator. This approach results in an allocation of training tasks to either the real system, or some kind of simulation. Whether a task is feasible for low-cost simulation or can only be trained with a full-blown simulator can be determined in light of the following criteria:

- Using a simulator for training should have an added value above training on the real system.
- A considerable part of training can be done with a simulator (at relatively low costs).
- The remaining training tasks (if any) can only be simulated at high cost.
- Those (remaining) tasks can safely be trained on the real system.

In a European project called ELSTAR[17] (Korteling, Helsdingen, & von Baeyer, 2000), an analysis was done to determine the feasibility for low-cost simulation of six main (military) domains: Maneuvering, Intelligence, Target acquisition and weapon delivery, Combat service support, Command & Control, and Maintaining mobility and survivability. Each of these domains was subdivided as far as it was affected by sub functions, organizational level, level of threat, and task or environmental conditions. Consider as an example the domain Maneuvering that has two sub functions: moving and navigating. In terms of the taxonomy moving can be done with wheeled vehicles, tracked vehicles, fixed wing aircraft, rotary wing aircraft, or a sea-platform. For each category in each domain, a list of fifteen aspects regarding training need, simulation need, knowledge generation, and simulation simplicity was checked. Based on the high score it yielded the sub domain 'wheeled vehicles' (i.e., cars or trucks) was appointed as very feasible for low-cost simulation and it was selected to build a demonstrator. In the army, many soldiers have to be trained for the particular task (training need), using the real equipment is costly and not optimally efficient (simulation need), the knowledge required for simulation was available in commercial of the shelf (COTS) technology (simulation simplicity) although not yet present in a satisfying simulator solution (generation of knowledge).

To make the low-cost demonstrator optimally efficient, Van Winsum and Korteling (1998) performed an extensive yet concise analysis of the driving task. They distinguished 17 so called 'elementary driving tasks' (EDTs). These EDTs were formulated relatively independent of each other (no redundancy). For each EDT they

[17] European Low-cost Simulation Technology for the ARmed forces

specified the relations with other objects and the required behavior, the applicable traffic rules, a summary of critical environment- and task-variables, as well as the perceptual motor, cognitive and procedural operations involved. Kappé & Korteling (1998) subsequently derived the requirements for a simulator from each EDT.

Based on the analyses done in the ELSTAR project, TNO was able to build a relatively simple simulator that nevertheless allowed basic training of about 60% of all driving tasks. It was estimated that the costs of training the remaining 40% would add a factor 30 to 300 to the cost of the simulator.

In order to keep the simulator really low cost, it was not able to provide motion information and the visual field was limited to about 180°. Tasks depending heavily on motion cues, for example, special maneuvering, terrain driving, high speed vehicle handling, and skidding could therefore not be trained in the simulator. Many other tasks, however, can be trained without provision of motion information. The same is true for the field of view. The number of tasks that require a person to have a direct view behind the car (i.e. 360° FOV) is rather limited. To generate such a large visual field would require a lot of effort that only facilitates a small part of training.

The low cost simulator (or LOCS as it was called) was built as a demonstrator and an environment to facilitate research into low-cost (driving) simulation rather than as a tool to provide training. In the year 2001 a commercial follow-up resulted in a successful part task simulator that was actually used in the training course of one of the largest driving schools in the Netherlands (Kappé, Van Winsum, & Van Wolffelaar, 2002). The experiments in this thesis were run on the original 'LOCS' demonstrator / research facilities. A description of this system follows in the next section.

4.5 The Low Cost Simulator (LOCS)

The Low Cost Simulator (LOCS) has been used for a number of experiments with regard to driver training. Three of these, explicitly related to instruction are reported in this thesis.

Figure 4. *The Low Cost Driving Simulator (LOCS)*

The LOCS was developed at TNO-Human Factors (TNO-HF) as a tool for research into simulator training, validity, transfer, and training effectiveness[18]. Besides, it also served as a demonstrator of the capabilities of low cost technology in training.

Each of the LOCS components will be briefly described below. *Figure 4* shows a picture of the LOCS. A schematic representation of the LOCS with its associated computer systems can be viewed in *Figure 5*. (See Appendix A for detailed technical specifications.)

The LOCS mock-up provides the interface with the simulation. It consists of a car seat, steering wheel, pedals (brake, acceleration, clutch), gear lever, and an (analogue) speed indicator. All components are original car parts (although from different cars) except for the speedometer, which is self made.

Furthermore, the LOCS consists of a number of computers that all serve specific purposes. Three Windows NT machines generate the visual environment on five wide screen (24”) displays that are positioned in a semi-circular configuration. Each of these computer monitors has a horizontal field of view (FOV) of approximately 40° resulting in a total FOV of about 200°. The central screen displays a high-quality image, which is generated by a single computer. The other two computers each generate a reduced resolution image on two peripheral displays. For that reason, the video-card of the central image generator is more powerful than that of the two other machines.

[18] Note that it is not a training simulator, that is, it was not designed to support transfer of (driving) skills.

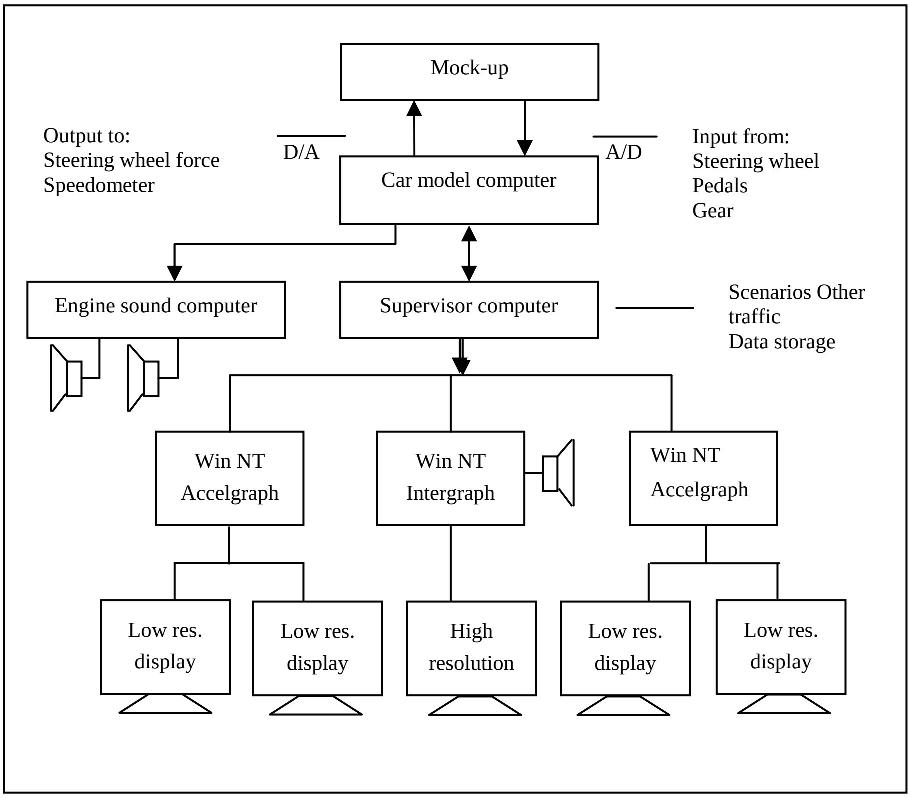


Figure 5. *Schematic overview of the components of the LOCS. More detail is provided in the text.*

Furthermore, one MS-DOS computer ('model computer') receives the signals generated in the mock-up (by standard A/D, DAC, and RS232 interfaces) and uses them to compute the corresponding vehicle behavior mathematical model. This machine provides input to the 'sound computer' (Windows ’98), which generates the appropriate engine sounds, and communicates with the 'supervisor computer'. The supervisor is used for scenario control, generation of other traffic, and data storage. The communication between these different computers is established through an Ethernet connection.

4.6 Database

Before the first experiment, specifications for a database were drawn up. These involved the layout of the road network, width and type of roads and curves, the

different types of intersections, and placing of the road markings, signs, trees, and houses in the database.
As can be seen in Figure 6 the database covers an area of 2.5 by 1.8 kilometers. There are two types of roads: 'rural' 80 km. roads (indicated by the dotted lines) and 50 km. roads (continuous lines). Houses (represented as blocks) are placed around intersections and in the center of the database, an apartment block is placed. The specifications were made up in accordance with the Dutch regulations, (RVV, 1990). The database was built by Mirage 3D in VEGA using MultiGen Creator.
The X in the upper right quadrant represents the starting point for a scenario in phase 2b (see also Appendix B).

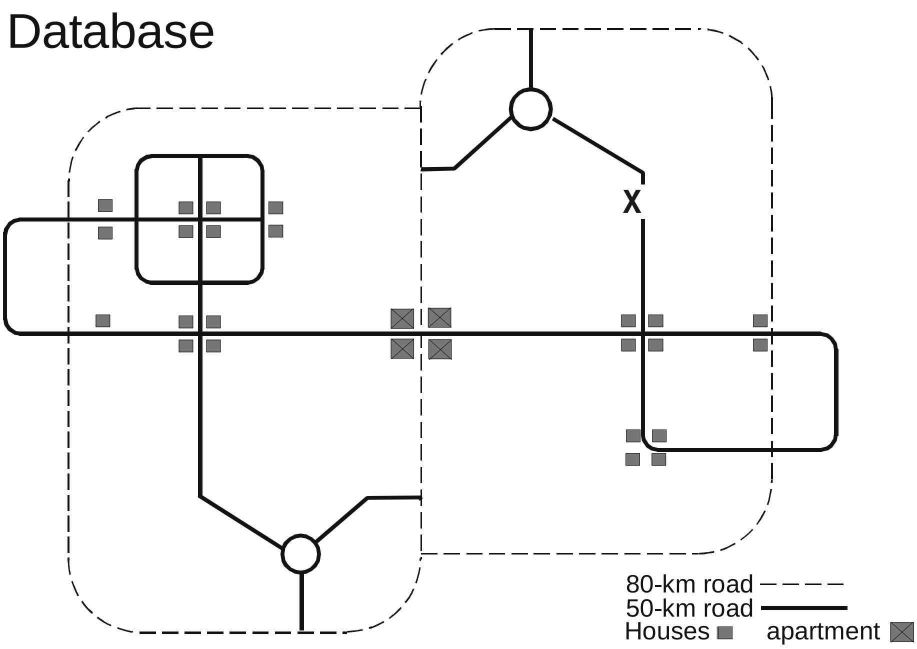


Figure 6. *The LOCS database plan.*

As will be explained in Chapters 5 and 6, the experiments consisted of phases. In each phase the trainee was confronted with additional tasks or tasks of higher difficulty. For these different phases of the experiments a number of different routes through the database were programmed. The scenario that is described in Appendix B has a prototypical route for phases 1a to 2b (these phases involved basic vehicle handling at low speed). The third phase involved driving 80 km/h roads only (these are the dotted lines in Figure 6). Finally, in phase 4, the complete database was used. Subjects received instructions which road to follow by means of a voice recording (saying 'turn left' or 'turn right') that was played at 100 meters before each intersection that required turning.

The low-cost simulator as described above combines science with computer games. As such, it might be a valuable tool for doing research on instruction (a field that itself has been looking for the balance between fun and utility for such a long time). Without pretending to be a completely valid (high-fidelity) representation of a real car, the LOCS has many important characteristics to train a task, as complex as car driving. This sets out the research approach for the experiments to come: at first exploratory (what characterizes the instructional process of a high-performance task in a simulator) and later on focusing more and more on those aspects of instruction that can make a simulator a successful didactical instrument.

Chapter 5

Exploration of Instruction

Abstract
This chapter presents a registration and analysis of the instruction process in a driving simulator. It was shown that a limited set of messages could be used to provide 70% of the instruction and feedback throughout the experiment. Furthermore, the overall judgment of a human instructor was fairly accurately predictable using six performance measures although an attempt to predict the instructions themselves (during practice) was only partly successful. It appeared that that there simply was no one-to-one mapping of instructions to errors. Apparently, to prevent excessive workload, human instructors leave many mistakes without consequence. Also, it was concluded that the instructor for a large part relied on the self correcting abilities of the trainees. Trainees liked participating and learned well during the relatively short period of time available in the experiment. This supports the idea that a driving simulator can be a valuable instructional tool.

5.1 Introduction

The present experiment focused on mapping the different instruction during the execution of training activities. The main reason to focus on this so-called tutoring phase is our conviction that tutoring is the most dynamic and time-critical aspect of instruction. Therefore, we expect workload for both trainee and instructor to be relatively high during tutoring. Tutoring is largely an intuitive process that is still poorly understood. The present experiment was undertaken to obtain a better insight into this process, expecting that knowledge of effective and efficient tutoring will help developers of training to make better use of the training and instruction factors their simulator offers.

This experiment was not set up to test a specific hypothesis. Because of all the interdependencies between training and instruction factors a more explorative approach guided by the following questions was chosen:

- What kinds of instructions and feedback does an instructor give during training of a vehicle control task?
- Is it possible to relate instructions and feedback (emanating from subjective instructor criteria) to (objective) vehicle parameters such as lateral position on the road, speed, etc…?
- If such a relation can be found, does it change for different (sub) tasks? For example, the presence or absence of other traffic could affect the relation between the lateral position of the trainee and instructions / feedback related to lane keeping.

5.2 Method

5.2.1 Subjects

Twelve subjects (7 males, 5 females) were recruited from the 'TNO subject data bank'[19] to participate in this study. Even though they were selected as subjects without driving experience three subjects nevertheless reported they had driven small distances in a real car once before (2 subjects) or occasionally (1 subject). Two subjects stated that they spent over 1 hour each day playing computer games that involved driving (racing), however, they were accustomed to the computer keyboard as an input device. The other participants reported to play computer games only occasionally.

Age of the subjects varied between 17 and 27 years (mean 21.3). Most subjects (n = 7) had no concrete plans to start taking driving lessons on short term (within the next 6 months). Of the others, only one had taken the theoretical part of the exam. All subjects wanted to get their driving license eventually. Subjects were paid for their participation and were all highly motivated.

[19] A large dynamic list with personal data of over 500 volunteers. New volunteers regularly are recruited by means of newspaper ads and verbal advertizing.

5.2.2 Experimental task

The experiment consisted of four training phases (each consisting of two parts) gradually increasing in difficulty (see *Table 2*). In the first part of each training phase, the trainee started practice in a world without other traffic (part A). When proficiency was rated sufficiently high in a consecutive test, other cars were introduced (part B).

During the first training phase of the experiment elementary driving skills were trained: steering and speed control (acceleration and braking). Subjects were instructed to drive at a steady speed of 30 km/h (approx. 20 mph) whenever possible. Just before a curve, speed was to be decreased to enable smooth turning.

In the second training phase, maximum allowed speed was increased to 50 km/h (approx. 30 mph). Subjects had to learn to use the gear lever to shift between first and second gears and neutral. (For simplicity's sake it was decided not to use the clutch-pedal throughout the experiment).

Consequently, in the next training phase the third and fourth gears were introduced as the maximum speed was increased to 80 km/h (approx. 50 mph).

In the fourth and final training phase extra difficulty arose from the many accelerations and decelerations (and thus gear changes) necessary to complete the trials. Furthermore, two roundabouts were added in this phase.

Table 2*. Schematic representation of the experimental task.*

Training phase	Trial	Instructor task	Pass?	Fail?
1A no traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 1B	Return to practice 1A
1B traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 2A	return to practice 1B
2A no traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 2B	return to practice 2A
2B traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 3A	return to practice 2B
3A no traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 3B	return to practice 3A
3B traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 4A	return to practice 3B
4A no traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	go to 4B	return to practice 4A
4B traffic	Practice	Instruction	n.a.	n.a.
	Test	Rating	finish	return to practice 4B

Note that during practice (see 'Trial' column) no decision with regard to passing or failing was made, hence the n.a. (not applicable) in the 'Pass?' and 'Fail?' columns

Each training phase (part) consisted of a practice and a test trial. During a practice scenario the subjects received verbal instructions from a qualified driver instructor. After completion of the scenario, a test run of comparable difficulty was started during which no instruction or feedback was provided. After each test the instructor decided whether the trainee was ready to transfer to the next level of difficulty or needed more practice at the same level of difficulty. The experiment ended when the trainee had completed the test trial of training phase 4b or when the available time was exceeded (see below).

In each training phase, two to four scenarios were available. Each scenario within a certain training phase was of about equal difficulty with regard to the number of left and right turns, the number of intersections, the density of traffic, and the situations that occurred. (A detailed description of a scenario appears in Appendix B). Whenever the instructor judged trainee performance on a test to be insufficient to continue to a higher level, one of the other practice scenarios (at the same level) was started.

A practice scenario took about seven minutes to complete, depending on a subject's driving style and skill. The test scenarios could be completed in about 4 minutes. Including a couple of minutes for starting a scenario and briefing a trainee, minimum time for completing a part of a training phase was 15 minutes. After an hour a standard 15 minutes break was inserted and subjects could also ask for a short break at any time during the experiment. A few times such extra breaks were considered necessary by the instructor. Without extra breaks, the minimum duration of the experiment was 2.15 hours. Once extra breaks and extra practice trials were required, the experiment would take longer. It was decided to set a maximum time for the experiment of 3.15 hours. We expected that (due to long duration with lack of progress) subjects' motivation would drop whenever they needed more than four extra trials. An additional benefit of this was that we could schedule two subjects each day (one in the morning and one in the afternoon). Subjects were very enthusiastic throughout the experiment and wanted to continue even during the break(s). Nevertheless a number of subjects (4) failed to complete all four experimental phases. In two cases the instructor felt that it was useless to continue the experiment because the trainees could not concentrate on the task anymore although they were still highly motivated (this was after more than -the minimum of- 2.15 hours).

5.2.3 Procedure

Subjects were informed they had to participate in an experiment that would take approximately three hours and was about training in a driving simulator. After being introduced to the instructor, subjects first had to fill out a small questionnaire with questions about their participation in traffic, experience with computer games, and knowledge of traffic rules. The complete form is included in Appendix C. The final question was a short test in which pictures of the seven traffic signs that were used in the simulator database had to be recognized. Incorrect answers were always followed by a correction and sometimes by a short extra explanation.

After that, subjects were invited to take place in the simulator and were helped to adjust the seat. Prior to each phase the functioning of relevant controls was explained and subjects received some general instructions. Subjects were told explicitly to keep to the speed limits in each phase. A practice scenario was always preceded by a short briefing on the task and the circumstances (presence or absence of other traffic, maximum speed, etc…). After the trainee completed a practice scenario, it was left to the instructor to provide some remarks about performance as additional feedback. During practice subjects were allowed to ask questions whenever something was not clear. In the test scenarios, they were told they would not receive any help unless the scenario could not be finished otherwise (these cases invariably meant that the subject 'failed' the test).

Figure 7. *Approaching an intersection in the LOCS as seen on the central screen from the drivers' perspective*

5.2.4 Variables and analysis

During the experiment, a number of vehicle parameters and instructor utterances were stored. With regard to these variables a distinction has to be made between practice and test trials. During the practice trials emphasis was on the instructional *process* whereas during the test trials the *product* of instruction (i.e., performance) was most important.

5.2.4.1 Practice trials

In the practice trials, (verbal) instruction and feedback was given in free format and recorded digitally. The vehicle parameters that were stored included the vehicle's lateral position, speed, and turn rate[20] (see *Figure 8*). These were sampled at 60 Hz. For the analysis of the practice trials the data files were divided in segments of 5 seconds, which could contain a single instruction. An instruction or feedback utterance was always placed in the segment in which it was started. For each segment (containing 5 * 60 = 300 samples) a mean value and a standard deviation was calculated for the three vehicle parameters. This way, two variables were derived from each vehicle parameter resulting in six variables (see *Table 3* for a summary).

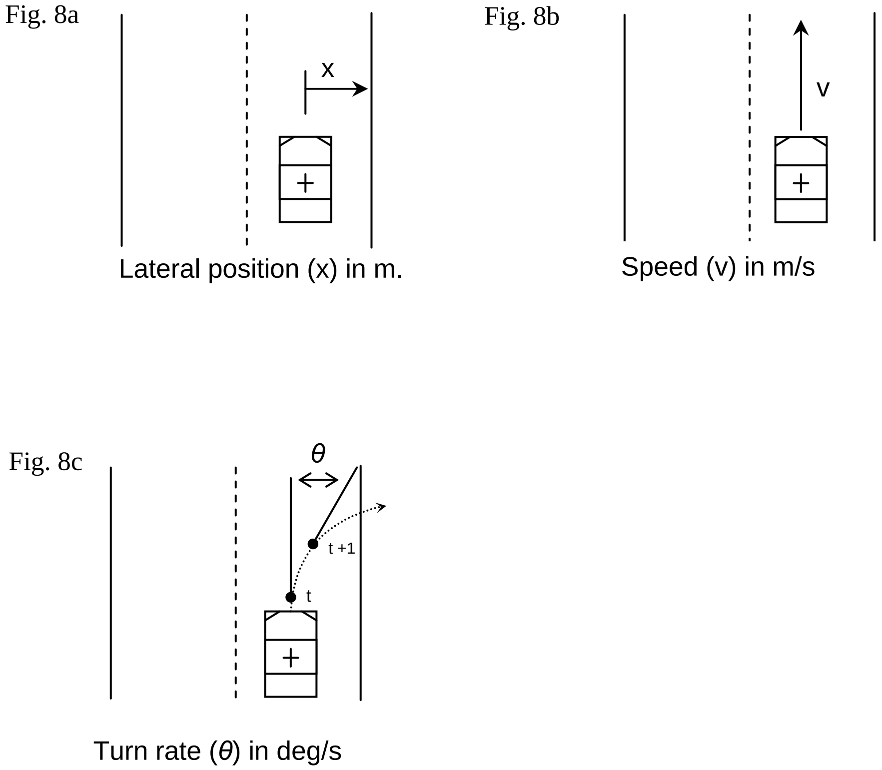


Figure 8. *The vehicle parameters measured in the experiment. Figure 8a: Lateral position (x) of the car in meters from right side of the road. The '+' on the car is the reference point used for measurement. Figure 8b: Speed (v) in kilometers per hour. Figure 8c: Turn rate (θ) in degrees per second.*

[20] The variable 'turn rate' is a derivative from the heading of the car. Heading itself could not be used in the analyses because its interpretation is troubled by inherent problems. The coordinate system used in the experiment was ground based. Straight driving could therefore occur at 0, 90, 180, or 270 degrees making it impossible to calculate a meaningful mean or standard deviation. Furthermore, the fact that a heading of 360° is equal to 0° results in strange artifacts in the means and standard deviations.

5.2.4.2 Test trials

In the test trials the instructor gave a single statement rating overall performance of the trainee at the end of the trial. Performance was judged either sufficient or insufficient to pass to a next phase of the experiment (a subject could 'pass' or 'fail'). For each three samples, the mean value and standard deviation of these parameters was stored (resulting in 20 means and standard deviations per second). The same variables were derived from the vehicle parameters as in the test trials (see *Figure 8* and *Table 3*).

5.3 Analysis

The verbal protocols of the instructor were scored for each practice trial. Only instructions and feedback directly aimed at performance were categorized. Motivational comments or more elaborate explications of a situation were left out of consideration.

In each phase of training, the amount of instruction and feedback was calculated as a percentage of total time. The data were inspected visually. It was decided not to use statistical tests because the tasks were not comparable *between* training phases as a consequence of task differences.

Within each phase a number of discriminant function analyses were conducted. Discriminant function analysis (DFA) is a mathematical technique similar to MANOVA in that it is used to compare differences among means of different groups relative to the overlap in their sampling distributions. Differences are evaluated as a ratio of variance between groups and variance within groups (error variance) and are tested for significance against a critical value of the F distribution.

The purpose of DFA is to predict group membership from a set of predictors. In a process called classification, all cases are assigned to groups based on their scores on the predictor variables. For each group a classification equation can be formulated by assigning weights to the predictor variables. The accuracy of the classification process is expressed in a classification matrix that shows the percentages of correctly and incorrectly classified cases.

Table 3. *The variables that were used for the different analyses*

Variables	Practice (1 value / 5 sec.)	Test (1 value / 0.05 sec.)	Unit
Predictor	Mean lateral position	Mean lateral position	m.
	SD lateral position	SD lateral position	
	Mean speed	Mean speed	Km/h
	SD speed	SD speed	
	Mean turn rate	Mean turn rate	Deg/s
	SD turn rate	SD turn rate	
Grouping	Instruction (free format)	Instructor judgment	pass/fail

As can be seen in Table 3, in this experiment six predictor variables were used to predict instructor judgment (test files) or instruction (practice files).

Because differences in performance attributable to road type would obscure the expected statistical effects, straight road segments and curves were analyzed separately.

In the analysis of curved road segments 50 meters straight road before and after such a segment were included. Consequently, the analysis of straight road segments excluded the first and last 50 meters of each segment.
Not every type of instruction was used with the same frequency in each phase or on each road type. From criteria with regard to the ratio between observations and predictor variables (Tabachnick & Fidell, 1989) it was determined that at least seven observations were necessary for an instruction to be included in DFA.

In phases 1b, 3b, 4a, and 4b the number of instructions on the straight roads as well as in the curves was insufficient to meet this criterion. Therefore, no analysis was done in these phases. Analysis of the left and right curves was possible only in phase 1a. Phases 2a, 2b, and 3a had enough samples to analyze instruction on the straight roads only. In Table 4 this is summarized: cells marked with a '+' represent phases that were statistically analyzed. Cells marked with a '-' represent phases in which insufficient numbers of instruction for statistical analysis were observed.

Table 4*. Possibilities for analysis.*

(practice) Phase	1a	1b	2a	2b	3a	3b	4a	4b
Straight road	+	-	+	+	+	-	-	-
Curve left	+	-	-	-	-	-	-	-
Curve right	+	-	-	-	-	-	-	-

In a similar way, DFA was applied to the test trials: If differences between subjects who 'failed' or 'passed' according to the instructor are reflected by the objective performance measures it should be possible to reliably discriminate between those two categories. Because it may again be expected that differences in performance that can be attributed to the different road types may confound the classification process, separate analyses were conducted for each road type (straight road, curve right, or curve left) in each phase. In phase 4b no analyses could be done because all subjects passed the test.

5.4 Results

Four subjects were unable to complete the full experiment in the available time. Two of these subjects still showed (a little) progression in the last trial before the experiment was ended. The other two seemed to get somewhat frustrated as a result of their lack of progress. Towards the end of the experiment they experienced problems concentrating as a result of which their performance deteriorated. This was considered no reason to exclude these subjects from the analysis because their data could be interesting, in particular from an instruction point of view.

5.4.1 Practice trials

A total number of 1459 instructions were given throughout the experiment. These instructions were divided into 38 different categories, each of which could be assigned to one of four main categories: lateral position, speed, gear, and interaction. The latter main category included instructions with regard to interactions and conflicts with other traffic and the relevant traffic rules pertaining to such interactions (right of way rules).

To estimate reliability of the scoring process, one trial was picked for re-scoring: once by the original observer and once by another person. The percentages of inter- and intra observer reliability were calculated for this trial by dividing the number of corresponding observations by the total number of cases. The correspondence with the original scoring was 96% on both occasions. This indicates that the instructions were unambiguous and scoring has been accurate. No other trials were selected to calculate inter observer reliability considering the excessive amount of time required for this.

The following figure (*Figure 9*) shows the change in amount of instruction over the experiment. Although no statistical tests could be performed to estimate significance, a general trend showing a decrease in amount of instruction towards the end of the experiment is clearly visible. The separate categories of instruction show a somewhat more complicated pattern.

Instructions in the main category 'position' are the only ones that show a general decrease (with the exception of phase 2b, which shows a slight increase). Main categories 'speed' and 'gear' (the latter is introduced in 2a) show a decrease within phases (part a vs. part b). Between phases, on the contrary, there does not seem to be a decrease until the final (fourth) phase.

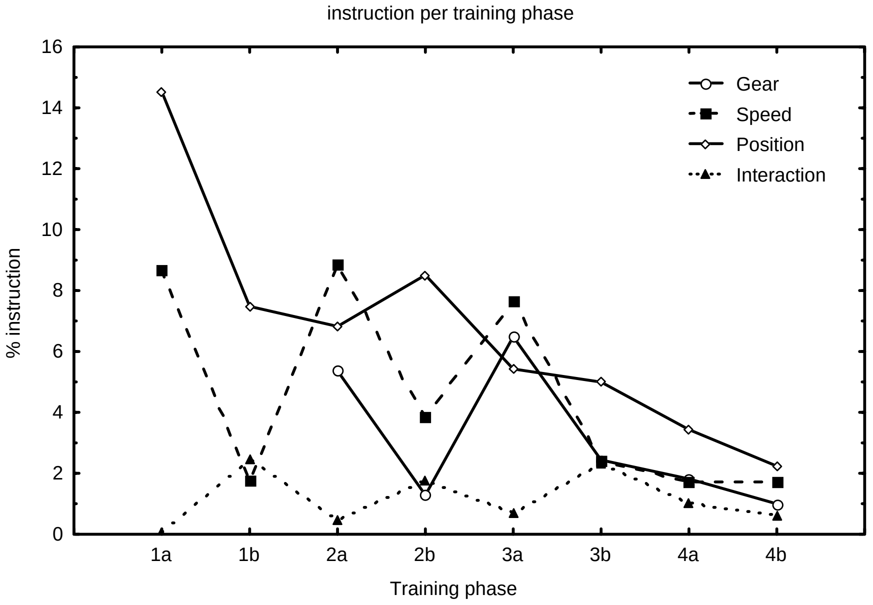


Figure 9. *Instruction shown as a percentage of total time.*

The percentage instruction in the main category 'interaction' was low throughout the experiment. Here, the addition of traffic (part b of each phase) seems to cause a small increase in instruction except for the fourth phase.
Table 5 provides a closer look at the instruction categories. It shows that in each phase only a limited number of instructions were given often enough to be included in the analyses (the table is continued on the next page). The instructions are placed in 38 categories. Instructions that were observed at least seven times in a particular phase are printed in bold face. It should be noted that the number of trials was not equal for each phase. Therefore, the decrease in instruction that seems noticeable in the table should be seen relative to the number of samples as in *Figure 9*.

Table 5. *The absolute number of instructions, per phase, and category of instruction*

code	*Content of instruction*	*1.a*	*1.b*	*2.a*	*2.b*	*3.a*	*3.b*	*4.a*	*4.b*	*sum*
10	Misc. position	**60**	**14**	**20**	**18**	**16**	**20**	**11**	5	
11	Position to left	**23**	**21**	**10**	**17**	**9**	1	4	2	
12	Position to right	**20**	3	3	4	4	1	3	1	
13	Position danger	1	2		3	2	1			
14	Position + speed	4	4	3	2	1	3	1		
20	Misc. turning	1	2	2		1				
21	Turn too sharp	**10**	5	**7**	**10**		1	4	5	
22	Turn too wide	**11**	3	5	3			3		
23	Start turning	**38**	**7**	**10**	**7**	2	2	2		
24	Faster steering	**46**	**17**	**12**	4	2	2	2	1	
25	Change position of hands	**14**	4	1					2	
26	Turn too early	**18**	**10**	5	6	1	2	3		
27	Turn too late	4	2	2	2	2				
28	Gentle steering	**27**	**12**	4	**10**	6	**8**	1	2	
	Position (total)	*277*	*106*	*84*	*86*	*46*	*41*	*34*	*18*	*692*
30	Misc. speed	4	1	**20**						
31	Accelerate	**22**	1	**10**	**13**	**19**	2	3	5	
32	Steady speed	4	1	**26**	1	**9**	3	1		
33	Too fast	**58**	**10**	**25**	**12**	**15**	5	6	4	
34	Brake	**53**	**7**	**8**	2	**11**	6	3	1	
35	Stop			1		1		1		
36	Speed in curve	**17**	4	4	1		2	2		
37	Speed + gear	**7**	1	**15**	**10**	**10**	2	1	4	
	Speed (total)	*165*	*25*	*109*	*39*	*65*	*20*	*17*	*14*	*454*
40	Misc. gear			1		1		1		
41	Gear up			**31**	5	**32**	6	6	2	
42	Gear down			**16**	3	**9**	5	3		
43	Gear + turn				2	2				
44	Gear earlier			1				1	2	
45	Gear later			7		2	2		3	
46	Gear problem			**8**	3	**8**	6	6	1	
47	Gear + braking			2		1	1	1		
	Gear (total)			*66*	*13*	*55*	*20*	*18*	*8*	*180*
50	Misc. interaction		**8**	2						
51	Look right				1					
52	Look left	1	**14**	4	**8**	2	5	1	3	
53	Check intersection		**7**		1	1	6			
54	Cutting in on another car				2		3			
55	Right of way		5		4	3	5		1	
56	Stop sign							**9**		
57	Collision		1		2				1	

Table 5. continued

	Interaction (total)	*1*	*35*	*6*	*18*	*6*	*19*	*10*	*5*	*100*
0	*Car not controlled*	3	1	4	8	3	4	5	5	33
99	*Missing data*	68	4	50	25	2	12	3	109	273
-10	*no instruction*	1392	1246	912	822	671	705	903	646	7297
	sum instruction	446	167	269	164	175	104	84	50	1459
	samples	1906	1417	1231	1011	848	821	990	805	9029
code	*Content of instruction*	*1.a*	*1.b*	*2.a*	*2.b*	*3.a*	*3.b*	*4.a*	*4.b*	*sum*

After compensating for the total number of instructions in each phase, the following instructions remained that were observed relatively often throughout the experiment. These referred to:

- Lateral vehicle position on the road in general;
- A position too far on the right side of the lane ('position to left');
- Steering too sharp into a curve ('turn too sharp')
- Starting point for steering in a curve ('start turning')
- Steering behavior while turning ('faster steering')
- Turning while still on straight road ('turn too early')
- Unsteady steering behavior in general ('gentle steering');
- Driving too slowly, ('accelerate');
- Exceeding speed limits ('too fast');
- Decelerating too little / or not in time ('brake')
- Inappropriate speed for gear change ('speed + gear');
- Change gears -up as well as down ('gear up', 'gear down');
- Looking for traffic while approaching an intersection ('look left')

5.4.1.1 Discriminant function analysis with regard to practice trials (by road type).

Analysis of the <u>straight roads</u> in phase 1a included the following instructions: -10, 10, 11, 12, 24, 28, 31, 33, and 34 (see Table 5 for explanation of the codes).

The result of the analysis is significant ($p<0.000$, Wilks’ Lambda = 0.621, approx. $F(48, 7197) = 15.262$). This means that classification of the instructions based on the predictor variables yields better results than randomly assigning the instructions to each sample.

However, inspection of the classification matrix learns that the number of cases predicted correctly is still rather low (less than 40% correct i.e., over 60% incorrect). Cases in which no instruction was observed during the experiment (-10) are often incorrectly classified (most often as instruction 34). Instruction number 33 (too fast) was also often incorrectly classified as instruction 34 (brake).

The confusions (incorrect classifications) between -10 and 34 show that no instructions were given (observed during the experiment) in many situations where instructions would be expected (predicted by the classification functions). Furthermore instructions 33 and 34 cannot be distinguished very well.

Analysis of the <u>right turns</u> in phase 1a included the following instructions: -10, 23, 24, and 36.

The result of the analysis shows a significant main effect ($p<0.0004$, Wilks' Lambda = 0.626, approx. F (18, 266) = 2.666).

The total classification process is only moderately successful (55% correct). Misclassifications occur for 'no instruction' with 23, and 36 (start turning, and speed in curve). Instruction 24 (faster steering), was often predicted where 23 (start turning) was observed. Again, the number of observed instructions was considerably lower than would be predicted.

Analysis of the <u>left turns</u> in phase 1a included the following instructions: -10, 23, 24, 33, and 34

Again, the result of the analysis is significant ($p<0.0001$, Wilks' Lambda = 0.576, approx. F (30, 482) = 2.373)

Inspection of the classification matrix learns that the number of cases predicted correctly is very low (less than 35% correct). Instruction 23 (start turning) appears to be responsible for this result as it is often incorrectly classified as 24 (turn back) and as 10 (misc. position). Furthermore, the classification often predicted instruction (any category) when in fact no instruction (-10) was observed during the experiment.

Analysis of straight roads in phase 2a included the following instructions: -10, 10, 11, 31, 33, 34, 37, 41 and 42.

In phase 2a only the straight roads could be analyzed. The results are significant ($p<0.000$, Wilks' Lambda = 0.758, approx. F (48, 4265) = 5.157). Correct classification, however, is only slightly over 36 %. Instruction 33 was often incorrectly classified as 34 a finding that replicates the results of phase 1a. Just as in the previous phase, cases were often classified in any category of instruction when in fact no instruction (-10) was observed during the experiment whereby instruction 41 was most often erroneously predicted.

Analysis of straight roads in phase 2b includes the following instructions: -10, 10, 11, 31, 33, 37 and 53.

Again the analysis results in highly significant values ($p<0.000$, Wilks' Lambda = 0.828, approx. F (36, 3133) = 3.814). The percentage of correctly classified cases is 39 %. From the classification matrix it was observed that cases were often classified as instructions 31 and 33 whereas 10 and 11 were observed respectively. Cases in which no instruction was given were often incorrectly classified as if instruction was given.

Analysis of straight roads in phase 3a includes the following instructions: -10, 10, 31, 33, and 41.

Again a highly significant result from the analysis ($p<0.000$, Wilks' Lambda = 0.748, approx. $F(24, 2049) = 7.422$). The percentage of correctly classified cases is relatively high (62 %). Although this seems promising, it still means that in 38% of the cases instruction was predicted incorrectly. From the classification matrix it was observed that misclassifications occurred mainly for instruction 31 incorrectly classified as 41, instruction 33 as 10, and –10 as any one of the instruction categories.

For all analyses it can be said that although the main effect was significant, due to the low percentage of correct classification inspection of the data in more detail was not justified.

5.4.2 Test trials

In the test phases, subjects could pass or fail a trial. A failure meant that a similar (not the same) practice and test trial had to be done. Only after passing for a test trial the subject could go on to the next training phase.
In *Table 6* the number of subjects that passed or failed the first test in each phase is displayed. Notice that due to a lost data file in phase 1a only data of 11 subjects were available, two subjects only made it as far as phase 2b, two more subjects dropped out after phase 3b.

Table 6*. The number of subjects in the categories 'passed' and 'failed' for each phase.*

phase	1a	1b	2a	2b	3a	3b	4a	4b
Subjects (n)	11	12	12	12	10	10	8	8
Passed	6	9	7	9	7	7	7	8
Failed	5	3	5	3	3	3	1	0

5.4.2.1 Discriminant function analysis with regard to test trials (by road type)

In (test) phase 1a, the results of the DFA were significant for the straight roads only ($p<0.000$, Wilks' Lambda = 0.658, approx. $F(6, 81) = 7.013$). 75% of the cases[21]

21 Each segment in the database was considered a case. In this test scenario there were eight (straight) road segments. The data of 11 subjects were stored resulting in 88 cases, 40 of which were from the 5 subjects that failed, the other 48 were from the 6 subjects that passed the test.

was classified correctly. The relevance of this percentage can only be judged when it is realized that the instructor judgment was given only once i.e., after finishing the trial. In fact it can be considered as a weighed sum of all observed cases during that trial. The classification of a case is based on the values of the performance measures in that case only. (In this analysis the curved road segments could not be analyzed because of the insignificant test results). So 75% of the cases could be classified correctly by looking at the straight road segments alone.

The classification-matrix shown in *Table 7* displays the correct classifications for both groups along the diagonal 'failed / failed' and 'passed / passed'. These two cells, printed in *italics*, can be called 'correct failures' and 'correct successes' respectively.

Table 7**. Classification-matrix of test phase 1a* **(straight roads***).** *The rows present the observed classifications*; *the columns present the predicted classifications.*

Classification-matrix of test phase 1a		*Predicted*			
		Failed p= 0.5	Passed p= 0.5	Total observed	Correct classifications
Observed	Failed	*22*	18	40	55.0 %
	Passed	4	*44*	48	91.7 %
	Total predicted	26	62	88	75.0 %

It can be seen that most of the 22 incorrect classifications (4 +18) were 'unjust successes' that is, the DFA predicted 'passed' whereas 'failed' was observed. Since each subject was measured on eight data points (or cases), misclassification will not inevitably result in a conflict with instructor judgment. A maximum of three misclassifications per subject was deemed acceptable here, as this would still leave five correct classifications.

On closer look, it can be seen that the incorrect classifications are concentrated on relatively few subjects. Subject 1 accounts for three of the four 'unjust failures'. Subjects 8, and 10 both have one case incorrectly classified.

Subjects 5, 9, and 11 showed 4, 5, and 8 misclassifications respectively. According to the instructor judgment they failed the test but the prediction (with regard to driving on the straight roads) classifies them as 'passed' based on performance in at least half of the cases.

The differences between the two groups of cases can be seen in the following figures: Figure 10 shows that the mean lateral positions per road segment in both categories ('passed' and 'failed') are roughly equivalent. The center of the car is always kept at about 2m from the right side of the road. For the subjects that failed the test however, the standard deviation of these means is about four times as large. This indicates that successful subjects kept a mean lateral position (per segment) closer to the mean (over all segments) than the subjects who failed the test.

Figure 11 confirms this finding for the standard deviation of lateral position per road segment. The mean standard deviation is smaller for the successful subjects. Furthermore the deviations from that mean standard deviation are considerably smaller for this group which indicates that these subjects were better able to keep the car steady than the subjects who failed.

All other vehicle parameters (except mean turning speed) show a similar picture: the successful subjects make fewer errors than failing subjects and if they do their error is smaller.

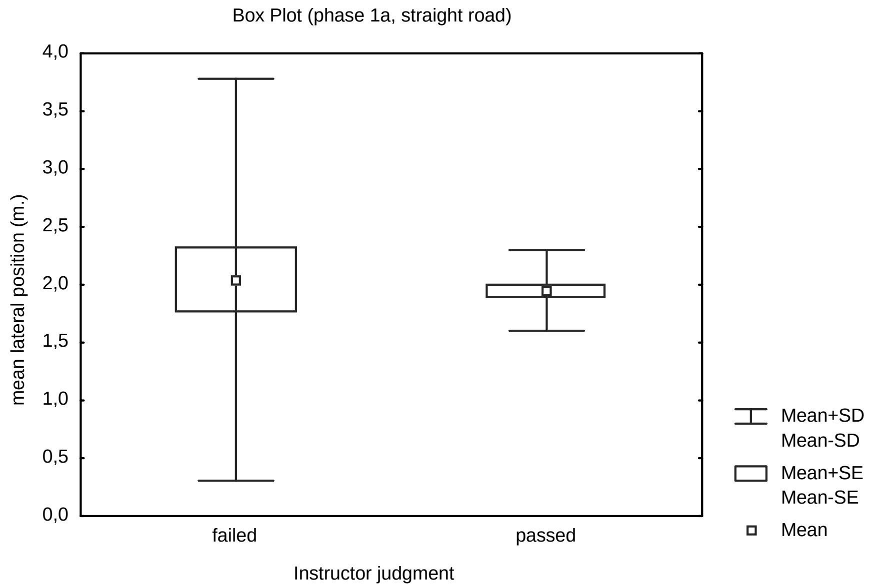


Figure 10 *Mean lateral position over the straight road segments in phase 1a.*

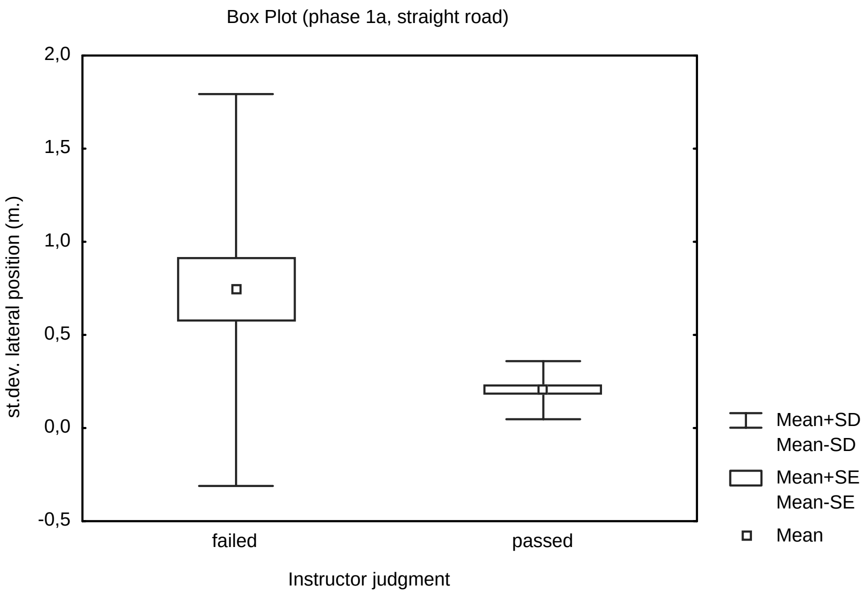


Figure 11. *Standard deviation of lateral position over the straight road segments in phase 1a.*

The results in test phase 1b were significant for all (3) road types.
For the straight roads ($p<0.0005$, Wilks' Lambda = 0.766, approx. F (6, 89) = 4.520) 76% of the cases was correctly classified. Misclassifications occurred mainly as incorrect failures (classified as 'failed' although the instructor judgment was 'passed')
For the right curves ($p<0.012$, Wilks' Lambda = 0.766, approx. F (6, 29) = 3.335) 89% of the cases was correctly classified.
For the left curves ($p<0.011$, Wilks' Lambda = 0.680, approx. F (6, 41) = 3.209) 79% of the cases was correctly classified.

When these data are combined, 16 classifications were made for each subject. A mismatch between classification and instructor judgment at the end of the trial would occur if eight or more cases (50% or more) were incorrectly classified. It can be seen that the maximum number of incorrect classifications for a subject (subject no. 5) was 7, which is equal to 44%. Therefore, it can be said that the occasional misclassification of cases did not conflict with overall instructor judgment.

In test phase 2a significant results were found for all road types. The percentages of correct classification were slightly lower than in phase 1b. This resulted in two (of 12) subjects with 50% (or more) incorrectly classified cases. In other words, for those two subjects the instructor and the DFA disagreed. For ease of reading, the statistical data are summarized in Appendix D.

Test phase 2b showed significant results for straight roads and left curves (see Appendix D for statistics). Combining the cases from these two road types results in 12 cases per subject. The data show that disagreement between instructor judgment and DFA classification would occur for subject 9, which had six conflicting classifications.

In the first part of phase 3 (3a) a significant result was observed for the right curves only[22]. The statistical data are again summarized in Appendix D. The incorrect classifications (10%) did not give rise to differences between instructor judgment and DFA classification.

In test phase 3b the test route was the same as in the previous phase except that it was driven from end to start, hence no right curves occurred. A significant result was observed only for the straight roads (see Appendix D). The relatively large number of incorrectly classified cases led to disagreement with regard to classification of three subjects (out of 10)

Test phase 4a combined the rural roads with the 50-km/h roads. In the analyses 4 road types were considered: Straight roads with a maximum speed of 80 km/h, and straight roads, right-, and left curves on the 50-km/h roads. The significance levels of the four categories can again be seen in Appendix D. Analysis of the rural roads showed only one case that was incorrectly classified. This did not lead to disagreement with regard to the overall judgment. For the other road types classification was perfect. No misclassifications were made; hence DFA never conflicted with instructor judgment.

No analysis could be done in phase 4b because all subjects passed the test.

5.5 Discussion and Conclusions

The present experiment has given an indication with regard to the nature of the instructions that are used during driver training and the frequency with which they occurred. Although it was possible to predict instructor judgment (at the end of a test trial) with a fair degree of accuracy (70 - 100%) using six performance measures, the attempt to predict the instructions themselves (during practice) was only partly successful. The percentage of correctly classified instructions varied between 30 and 40 % (with one exception of 65%) in the different phases of the experiment. Although this was significantly higher than chance level it must be concluded that this was not sufficient for accurate classification of samples in the appropriate instruction categories.

Several factors may have contributed to this:

[22] Bear in mind that the third phase involved driving on rural (80-km/h) roads. As opposed to the earlier phases, curves were designed for the maximum speed of 80 km/h. Furthermore, the trials in this part included no left curves.

- The large number of instructions that could be given and the way of presentation (free speech) could have obscured or confounded possible effects.
- The timing of instructions in relation to specific behaviors, or
- Other factors such as training history, or general impression of trainee skills

With regard to the first point it was found that in this experiment 70% of the instructions were scored in one of fourteen categories (a relatively small number considering the total number of 38 categories that was observed). Therefore the fact that free speech was used for instruction does not seem to be the major factor responsible for the disappointing accuracy of the classification.

A complicating factor, however, is that on average in only 16% of all samples instruction in one way or another was given. Since the amount of instruction decreased during the course of the experiment, the percentage of instruction in the last phases was considerably lower than 16%. Therefore, in phase 3 and 4 it turned out to be difficult to find enough samples in any category to perform statistical analyses. Furthermore, the delivery of instruction did not always follow consistently from the performance measures. In a large number of samples no instruction was given although the values of the performance measures suggested that some kind of instruction would be necessary. It is most likely that the instructor gave the subjects an opportunity to correct some of their mistakes before she intervened.

As far as the second point is considered (timing of instruction) the methods used in this experiment proved to be sub-optimal. It was necessary to score instructions within blocks of 5 seconds to be able to make comparisons between those blocks. This procedure introduced a timing bias and rendered it difficult to determine the exact start and end of an instruction. Furthermore, another source of timing bias resulted from difficulties with regard to the event that triggered a particular instruction. Instruction could refer to observed behavior or be pro-active (that is when no - faulty - behavior had occurred yet). Within this experiment it was not possible to correct for this problem

Finally, subjective factors may have played a part in deciding what instructions were given. This cannot simply be said to affect the effectiveness of instruction negatively but it does complicate research. From one point of view the (apparent) lack of consistency in the provision of feedback may slow down the learning of certain behaviors. On the other hand consistent provision of feedback would probably heighten the instructor workload to unacceptable levels. Furthermore, it may cause the trainee to become dependent of instruction but it might also interfere with the learning process. It may therefore be expected that an instructor tries to find an optimum between consistency of instruction and workload by including subjective factors such as training history and assumptions about trainee skills when deciding to give instruction or not.

Inconsistent delivery of instructions does not imply that the trainees’ mistakes are not noted. The fact that the instructor judgment at the end of the test trials could be predicted with high accuracy indicates that the overall judgment is constructed around a number of observations of performance. In this experiment, the instructor seemed to be very well capable of giving an overall judgment that was in line with objective performance measures although no criteria for instructions could be derived from the data during practice.

In connection with this observation, an experiment by van Rooij and Korving (1995) is worth mentioning. Using the Space Fortress Game (SFG) (Mané & Donchin, 1989) as experimental task, the authors designed six instructional interventions that could be presented to a trainee after completion of a single game. At first they asked an expert player (instructor) to select interventions for a considerable number of recorded games. With the DFA that was consequently applied to the data file they were able to predict the instructor judgment correctly in 86% of the games[23].

Their results showed a reduction of training time of 55% over a control group, which did not receive these instructions. Still the control group received the same briefing and followed the same training schedule as the experimental group. These results indicate that instructional interventions may be derived from performance measures (even though the instructions were not delivered during play).

In the practice trials, another interesting observation could be made with regard to the amount of instruction. It was seen that instructions considering lateral position showed a gradual decrease within as well as between phases. Categories 'speed', and 'gear' on the other hand show a pattern in which the decrease of instruction was much larger within than between phases. Since new tasks that were introduced referred to speed, gear shifting, or interaction with other traffic it appears that none of these tasks had an effect on the ability of trainees to control their lateral position.

Furthermore, it may be concluded (with some reservation) that the addition of a new task such as shifting gears imposed more of a burden than addition of traffic did. This should of course be seen in relation to the difficulty of scenario's in the simulator. The amount of traffic was limited yet enough to lead to a (small) rise in instruction with regard to interaction as can be seen in *Figure 9*.

Trainees clearly have learned something during this experiment in a relatively short period of time. By the end of the experiment, most of them were able to control the simulated car and drive through the database, on curves and on straight roads, at different speeds, in different gears, and interacting with other traffic. Whether training has been *efficient* cannot yet be determined. A follow up experiment will be conducted to compare the efficiency and effectiveness of different instructional strategies. It is suggested that this experiment be focused on simulator specific aspects of instruction. In particular when instruction is based on the principle of augmented cueing it might be more efficient as compared to verbal instruction.

This hypothesized benefit of using augmented cues as a means of instruction may be explained by the direct coupling it can have with the trainee's

23 This is comparable to the results of the DFA in the test trials of the present experiment. In these trials judgment was also given after a trial was completed. Only the type of judgment differed.

actions. When properly designed, augmentation immediately ‘shows’ the trainee what action to take. In line with this, Lintern and Roscoe (1980) state that:

“…. [It] might be possible to exploit the capability of [a simulator] to provide a learning experience different from that provided in an airplane. Some training manipulations that are not feasible in the air might moderate the trial and error process that is generally necessary to learn to integrate the correct perceptual motor responses for landing an airplane.”

Verbal instruction, as opposed to an augmented cue, always does require an extra step to translate the content of the message into action. Furthermore, speech is essentially serial in its nature, which means that the meaning of a speech sequence can only be extracted after some words have been uttered. E.g., when an instructor tells a trainee to go a little more to the right, the trainee won’t know what to do until the final word has been spoken.

Especially for continuous tasks such as steering and speed control, continuous feedback mechanisms such as augmentation might be beneficial because there exists a compatibility between the instruction and the appropriate response (see also Wickens, 1992).

Furthermore, as soon as more is known about the effects of instruction on performance in a simulator, questions relating to validity and transfer will need to be answered. Even if the learning process in a simulator turns out to be efficient and effective, the ultimate goal of driver training is not to drive a simulator but to drive a real car. This will be the ultimate yardstick to determine whether simulators have earned a place in driving instruction.

Chapter 6

Augmentation in a Driving Simulator

Abstract

The second and third experiments are reported in this chapter. Both experiments attempted to compare the efficiency and effectiveness of two different modes of instruction: the verbal instructions and a set of augmented cues. In experiment2, it was found that augmented cues were provided more often than verbal instructions throughout the experiment. Also, the number of subjects that was able to complete the experiment was smallest in the augmented condition. Three possible explanations are provided for these findings. 1). It might be that augmented instructions were experienced as less intrusive and thus were provided more easily. 2). Augmented instructions might be less effective than verbal instructions, or 3). A-priori skill differences between the two groups were responsible for the findings.

Experiment3 was set up to control for differences between subjects as much as possible by matching subjects on their aptitude for learning to drive. Again, it was found that the amount of instruction delivered to the augmented group was larger than to the verbal group. However, this difference was only significant in the first phase and strongly related to the low aptitude group thus indicating that the between group differences found in experiment2 were largely due to a-priori skill differences and not to the effectiveness of instruction. The data provide further evidence for the suggestion that augmented instruction is easier to provide than verbal instruction. If there is no need to provide additional instruction, it is used just as much as verbal instruction. This latter type of instruction seems to lack this flexibility. For low aptitude trainees then, it can be a solution to provide non-verbal (augmented) instructions because they seem to be friendlier.

6.1 Introduction

In the previous chapter it was shown that provision of verbal instruction and feedback during task performance was effective[24] in learning to drive a simulated car. Starting from scratch, the twelve subjects learnt to steer, accelerate, brake, shift gears, and interact with other traffic. Eight subjects reached the highest proficiency level of the experimental task in two and a half to three hours. Nevertheless, this experiment was exploratory. It was not designed to address the full didactic potential of the simulator. Instead, the instructor was asked to work up to a training session that was as similar to reality as possible (see Chapter 5).

Having in mind the didactical advantages of simulators compared to the operational environment it was decided to run a follow up experiment focusing on one aspect that seemed promising: Based on ideas and experimental findings from the literature (e.g., Lintern et al., 1987; Roscoe, 1991; Schneider, 1985) it was hypothesized that deviations from reality could enhance training effectiveness and efficiency. A concept that has been used in this connection is augmented cueing (O'Shea, Cook, & Young, 1999; Young, Stedmon, & Cook, 1999).

Augmented cueing works by emphasizing elements in the (virtual) environment to help trainees focus on the relevant characteristics of a task. This is particularly relevant during the initial phases of training. It helps trainees to understand the task faster which in turn helps speeding up the process of learning. Experimental research on augmented cueing has yielded mixed results, likely as a result from the multitude of aspects affecting instructional efficiency (Lintern & Koonce 1992; O'Shea, Cook, & Young, 1999).

Connecting this to the instructional process, augmented cues perfectly match to the previously mentioned skills cycle of Romiszowski (1999) (see Chapter 2). By providing a direct link between perception and performance, these cues will allow the trainee to make a shortcut through the skills cycle (directly from perception to performance). The 'recall' step (step 2 in the skills cycle) may be skipped because the augmented cues inherently present the prerequisite information. The appropriate reaction follows almost effortless. In this light it may be expected that instruction based on augmented cues is barely seen as intrusive. This is likely to yield improved performance compared to traditional (verbal) instruction.

The present chapter describes two experiments that were conducted to compare verbal instruction and feedback with instruction and feedback based on augmentation (these experiments will be referred to as experiment 2 and 3 respectively). Both experiments involved the between groups comparison of the two types of instruction and feedback.

To make a fair comparison, both types of instruction and feedback should be tested in the same environment. If one decided to test the verbal instruction in a real car and the augmented instructions in a simulator it would be plausible that a part of the results would have to be attributed to differences in the hardware (car vs. simulator). Which part that would be cannot be determined as Korteling and Sluimer

[24] no statements about efficiency were made.

(1999) point out. This problem was avoided in the present research by using the simulator in both conditions.

Another risk of misinterpretation arises from the differences between instructional content in the conditions. We were interested in the differences between two forms of instruction and feedback (verbal vs. augmented). If, additionally, differences in content were introduced, this would obscure the experimental results. A fair comparison, therefore, can only occur if the two conditions differ in form only. To assure this, the intentional message of the instructions and feedback was similar in both conditions.

In both conditions instructions and feedback were limited to the same set of actions / mistakes. Since the majority of the verbal instructions that were administered in the first experiment could be classified in relatively few categories, these were used to create the instruction for experiments 2 and 3.

To further minimize the possible disturbing effects of inadvertent differences in the instruction both the verbal and the augmented instructions were pre-designed and recorded and assigned to buttons. This way it was ensured that the content of a single instruction did not vary.

6.2 Experiment 2: Augmented instructions

6.2.1 Method

6.2.1.1 Instrumentation (the simulator)

The Low Cost Simulator (LOCS) has been developed at TNO-Human Factors (TNO-HF) as a tool for research into simulator training, validity, transfer, and training effectiveness. It also serves as a demonstrator of the possibilities of low-cost simulation. The basic configuration has been described in chapter 4 and Appendix A of this thesis. In the present chapter only the changes with respect to the basic configuration will receive attention (see also *Figure 17* and *Figure 18* further in this chapter).

Since the previous experiment a number of changes were implemented in the simulator (compare *Figure 5* in chapter 4 and *Figure 12* below). As a result the performance of the simulator was significantly improved. The graphical cards were upgraded and two extra PCs were added so that each screen was controlled by it’s own computer. This increased the update frequency of the images and guaranteed a smooth transition of one image to another. Furthermore, three LCD screens representing the left, right, and rear-view mirrors were attached to another image generator, and an LCD display was connected to the sound PC and generated the dashboard image (rev counter, speedometer, odometer, clock, gear indicator, and winking indicator). The gear indicator in the display was primarily meant for the instructor who often had no direct view of the gear lever. This new display presented a great improvement to the (simple) mechanical display that was used in the previous version of the LOCS.

6.2.1.2 Database

Besides the changes to the simulator hardware reported above, there were also changes related to software. The visual database, which generated the virtual environment, was enhanced and appeared to be more realistic (Appendix E). The layout (road net) remained basically unchanged from the previous experiment except for the roundabouts. These were enlarged and instead of having three-legs, they were changed to four-legged roundabouts. The reason for this was the fact that people had too much trouble driving the roundabouts in the previous version of the database. This change also required the creation of three new intersections in the vicinity of each roundabout. Furthermore, the number of trees and houses was increased considerably and traffic lights were added. *Figure 13* provides an overview of a typical intersection in the built up area of the database. The virtual environment hereby gained in attractiveness and realism. More important, subjects were required to look more carefully when approaching an intersection as some of the buildings obscured their view to the sides.

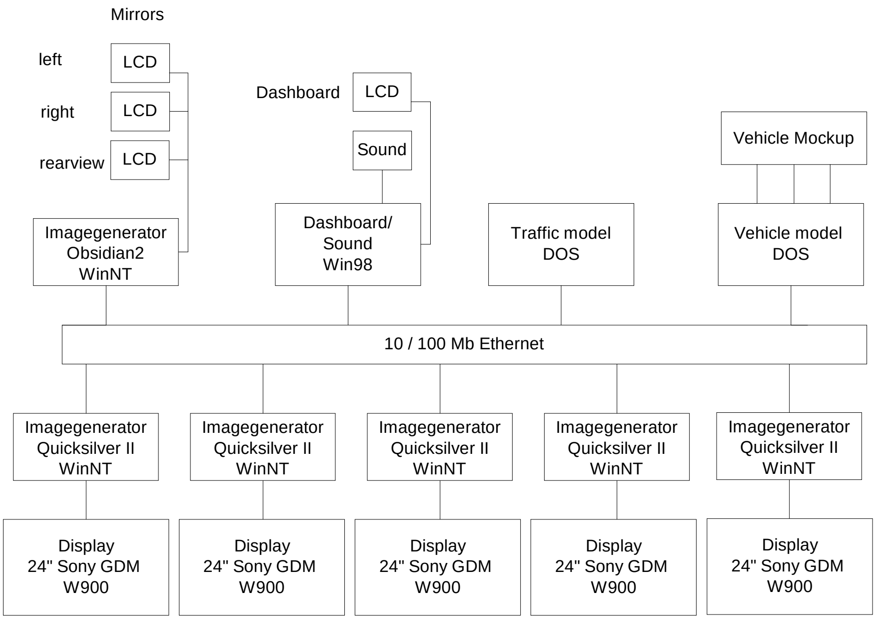


Figure 12. *Revised LOCS configuration (used in experiment 2 and 3)*

6.2.1.3 Subjects

A total of 33 subjects participated in this experiment. However, the data of only 32 subjects (11 male, 21 female subjects) were used for the statistical analyses because problems with the data storage resulted in a considerable loss of data for one subject. Because the remaining data files were not sufficient for analysis, these subject's files were excluded from the data set.

The average age of the subjects was 21.1 years with a standard deviation of 2.7 years (youngest 18 oldest 29). None of the subjects had received driving instruction in a car (or a simulated car) prior to the experiment. In an attempt to rule out a possible effect of sex differences care was taken to assign male and female subjects evenly to both conditions (in the verbal instruction condition the ratio male to female subjects was 6:10, in the augmented instruction condition it was 5:11).

Figure 13*. A typical intersection in the database (only the image of the central screen is provided)*

6.2.1.4 Experimental task

The experiment consisted of four phases in which the subjects were taught to drive in the simulator. Before each phase, the instructor explained the task and gave some general instruction. Afterwards, she provided additional feedback if deemed necessary.

Each phase introduced new aspects to the task resulting in a gradually increasing difficulty (see also *Table 8*). After a short introduction to the simulator the subjects were familiarized with steering and speed control (acceleration and braking) in the first phase. Subjects were instructed to drive in first gear only and act up to a speed limit of 30 km/h. Curves were to be taken at a slightly lower speed.

In the second phase autonomous cars were present. These traffic participants were not under direct control of the experimenter or instructor but either drove a random course through the database (to generate a flow of traffic) or a pre-programmed route (to create specific situations in traffic). The autonomous cars drove in accordance to the traffic rules with regard to right of way, maximum speed, overtaking, etc. In addition they possessed a certain kind of intelligence for example to prevent a 'lock up' when four cars arrive at an intersection at the same time.

Subjects were instructed to pay attention to the rules of the road and look carefully at each intersection before negotiating it. All other aspects of the task stayed the same as in the previous phase.

After this second phase, subjects learned to shift gears between first and second gear. The maximum speed was increased to 50 km/h. but no other traffic was present. In the fourth and final phase cars were added again.

Each phase comprised a practice trial and a test trial. Subjects received instructions and feedback in the practice trial only. After the test trial was finished, the instructor decided whether the subject was ready to go on to the next phase or needed more practice at the same level. Both the provision of instructions and feedback, and the final decision after the test were subjective (that is: by sound judgment of the instructor).

The aspects of the driving task that were taken into account were related to control of the car (lateral position and speed control, smoothness of steering, gear shifting), and insight in traffic rules (negotiating intersections, dealing with other traffic, abiding the rules). The instructor had been licensed as a professional driving instructor (for 6 years) and had previous experience with both driving and teaching in a simulator.

***Table 8**. Phases of the experiment.*

Training phase		Trial	Instructor task	Pass?	Fail?
1	no traffic + 30 km / h	Practice	Instruction	n.a.	n.a.
		Test	Rating	go to 2	Return to practice 1
2	traffic + 30 km / h	Practice	Instruction	n.a.	n.a.
		Test	Rating	go to 3	Return to practice 2
3	no traffic + 50 km / h	Practice	Instruction	n.a.	n.a.
		Test	Rating	go to 4	Return to practice 3
4	traffic + 50 km / h	Practice	Instruction	n.a.	n.a.
		Test	Rating	finish	return to practice 4

6.2.1.5 Conditions

There were two conditions that were administered in a between subjects design. In the 'verbal instruction condition' the subjects received verbal instruction and feedback in a preprogrammed format. For that means, twelve instructions were selected (see also *Table 9*). These utterances were recorded and stored on a personal computer (wave format) and assigned to the function keys of a standard PC keyboard (F1 to F12). Their meaning and relevance is explained in the text below.

The choice for these particular twelve utterances was largely inspired by the previous experiment (see Chapter 5). Here it was seen that a relatively small number of instructions was used regularly (in fact about 70% of the instructions that were provided could be done with a small set of 14 instructions). Based on this subset twelve instructions that could be presented in the preprogrammed format used in the present experiment were selected.

In the 'augmented cues condition' each of these twelve instructions or feedback interventions was translated into a non-verbal cue. This could be a visual, aural, or kinesthetic stimulus.

The visual instructions were traffic signs that were displayed on the central screen of the simulator after the instructor depressed the corresponding key. When the trainee had to adjust lateral position, a cue-line was displayed on the road contra-lateral to the side that was exceeded. This way the subject was made aware of his error. Subjects were instructed to steer towards the cue-line, which disappeared again after two seconds (see for example *Figure 14*).

Kinesthetic cues were delivered via the actuator that also provided normal steering wheel force. Such a cue consisted of a well noticeable jerk to the steering wheel to signal the moment of turning into a curve. The instructor pressed the corresponding function key at the moment appropriate for starting to turn left or right whenever she observed the subject having problems to estimate this moment correctly.

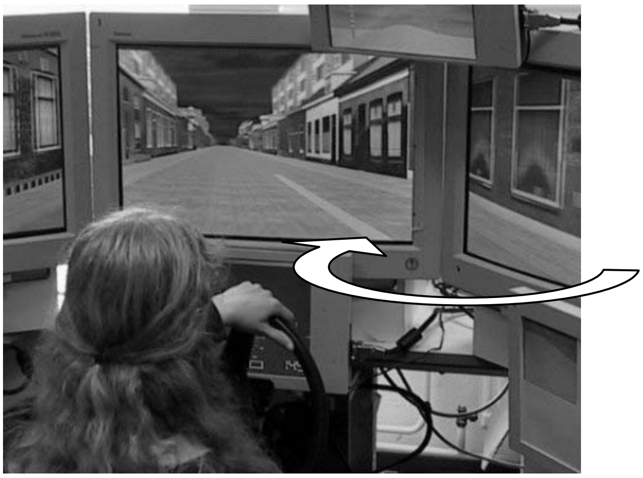


Figure 14. *The arrow points to an augmented instruction which is presented on the central screen of the LOCS. A bar (protruding from the virtual car) has appeared. It indicates a lateral position too far to the left; hence, the driver should steer right (towards the bar) to try to align the bar to the curb**.** It can be seen that the ideal lateral position of the car is with the curb aligned to the lower right corner of the windshield (i.e. the central screen).*

Finally, there were a number of aural cues. To alert a subject with regard to viewing behavior at intersections, three consecutive horn signals[25] could be played whenever the subject did not look for cars from left or right. Furthermore, when subjects drove too slowly the instructor could play the sound of a roaring engine (gradually going up in revs twice). To encourage a subject to change gears up or down, the sound of the engine was boosted with 2000 revs or lowered to 1000 revs respectively. This sound continued until the subject adjusted the gear position or the instructor ended the signal.

In both conditions the instructions were thus provided by means of keys presses. In the 'verbal condition' each key was linked to a pre recorded message. In the 'augmented condition' the keys were linked to a visual stimulus, a kinesthetic stimulus, or a sound. The decision to press a key at a certain moment during the experiment was always taken by the instructor.

[25] These sounds alternatively came from the left, right, and left, in accordance with the desired order of looking during the approach of an intersection.

***Table 9**. Semi automatic instruction: utterances or actions in each condition*

F-Key	Instruction Verbal (recorded voice message)[26]	Augmented
1	‘Come on, accelerate’	Roaring engine sound
2	‘Not faster than 30’	Traffic sign displayed on screen
3	‘Not faster than 50’	Traffic sign displayed on screen
4	‘Stop’	Traffic sign displayed on screen
5	‘Little bit left’	Position line displayed on the road
6	‘Little bit right’	Position line displayed on the road
7	‘Attention! Give right of way’	Traffic sign displayed on screen
8	‘Look carefully at the intersection’	Three beeps (left-right-left)
9	‘Start turning left now’	Jerk at steering wheel
10	‘Start turning right now’	Jerk at steering wheel
11	‘Gear up’	Constant high ‘revs’ sound
12	‘Gear down’	Constant low ‘revs’ sound

6.2.1.6 Routes
The experimental routes were mainly in the built up area. Appendix E shows the layout of the road network as it was created from the visual database. For each level in the experiment, two different routes were available.

6.2.1.7 Procedure
Subjects were informed they had to participate in an experiment that would take approximately three hours and was about training in a driving simulator. After being introduced to the instructor, subjects first had to fill out a small questionnaire with questions about their participation in traffic, experience with computer games, and knowledge of traffic rules. The complete form is included in Appendix C. The questionnaire ended with a short test in which pictures of the seven traffic signs that were used in the simulator database had to be recognized. The test was always followed by a correction of possible mistakes and if necessary by a short explanation.

After that, subjects were invited to take place in the simulator and were helped to adjust the seat. Prior to each phase the functioning of relevant controls was explained and subjects received some general instructions. Subjects were explicitly told to keep to the speed limits in each phase. A practice trial was always preceded by a short briefing on the task and the circumstances (presence or absence of other traffic, maximum speed, etc…). After a practice trial, the instructor sometimes made a remark about the performance. During practice subjects were allowed to ask questions whenever something was not clear. In the test trials subjects were not to receive any help unless the scenario could not be finished otherwise (these cases invariably meant that the subject ‘failed’ the test).

[26] Translated from Dutch. All messages were kept as short - and to the point- as possible (6 syllables at most in Dutch).

6.2.1.8 Variables

During the experiment, a number of vehicle parameters and instructor keystrokes (used to provide instructions and feedback) were stored. With regard to these variables, a distinction has to be made between practice and test trials. During the practice trials emphasis was on the instructional process whereas during the test trials performance was most important.

Performance was analyzed with regard to instructor judgment (the amount and nature of the messages they provided via the keyboard during practice, and the overall judgment during test), and trainee performance as measured on six dependent variables:

Mean lateral position, SD lateral position, Mean speed, SD speed, Mean turn rate, and SD turn rate. These were the same variables as in the previous experiment. The two instructional conditions (verbal vs. augmented) served as independent variables.

In the practice trials, instruction and feedback was given by means of the function keys. In the test trials no instruction or feedback was given. The instructor gave a single statement rating overall performance of the trainee at the end of the trial. Performance was judged either sufficient or insufficient to pass to a next phase of the experiment (a subject could ‘pass’ or ‘fail’). Furthermore, the same vehicle parameters as in the previous experiment were sampled: lateral position, speed, and turn rate. For each three samples, the mean value and standard deviation of these parameters was stored (resulting in 20 segments with means and standard deviations per second).

6.2.2 Hypotheses

For equal amounts of instruction in a practice trial, it was expected that the performance in the test trial would be better for the subjects in the augmented instruction condition. Good performance was defined as keeping close to the prescribed speed, with few sharp accelerations or decelerations. A lateral position of about 40 centimeters distance from the curb with a small standard deviation (lateral accelerations kept to a minimum), and finally a steady turn rate in the curves and at the intersections.

Based on experimental findings that inexperienced subjects profited most from augmentation, Lintern et al. (1987), concluded that augmented feedback would be useful at least for primary and intermediate instruction. Therefore, in the present experiment we expected that the difference between the augmented instruction group and the verbal instruction group (as it concerned the number of trials they needed to complete each phase) would be most pronounced in the first two phases of the experiment.

As it can be said that each new phase represents an increase in difficulty, it was expected that differences in the (over all) amount of instruction would be caused by those categories of instruction that reflected on the new aspects of a task. Initially, these would be speed and lateral position. In the second phase, instructions with regard to interaction with other traffic and traffic-rules were expected to dominate. Instruction in the third phase would reflect gear changing and speed control aspects, and finally an increase in instruction with regard to traffic rules was expected. The relative amount of instruction was expected to decrease towards the end of the experiment.

6.2.3 Analyses

6.2.3.1 Practice

Within the practice trials, for each phase, a t-test was used to determine whether the amount of instruction differed between the two instructional conditions. Because not every instruction (or feedback message) could be given in each phase, no quantitative statistical methods were used to make a comparison of the amount of instruction between phases. To provide at least some form of comparison, it was decided to provide a visual inspection of the data. We had to apply a transformation to the data to correct for differences in the number of trials per condition as well as for the different number of trials per phase. To be able to present the amount of instruction as a percentage of available time, each data file was divided in segments of five seconds in which either an instruction was started or not. After that the number of segments with instruction was divided by the number of segments without instruction.

6.2.3.2 Test

The test files were combined in a single data file, which was analyzed with three separate MANOVAs with the following factors: condition, road type, and -since there were no instructions provided during test- the factor phase was also included. Factor condition had two levels (augmented instruction and verbal instruction). Three different levels of road type were distinguished: straight roads, right turns and left turns. Finally, each phase represented a separate level for the factor phase.

However, because not all subjects participated in all phases one MANOVA was done on the first two phases (where all subjects participated), and the last two phases were tested with a separate analysis each. For each MANOVA we were interested mainly in the first factor (condition) and possible interactions of condition with road type or phase. The latter factors were expected to yield significant yet trivial main effects (e.g. it is obvious that on a straight road, driving speed will be higher than in curves). Therefore, comparing the values of the dependent variables on straight roads and curves or between the different phases was deemed irrelevant. These differences will therefore not receive further attention in the analysis.

With regard to the number of subjects that passed on to the next phase for each condition the data were visually inspected. No statistical analysis could be performed here because of differences in number of trials for each subject and the small number of subjects that made it to phases 3 and 4.

6.2.4 Results

6.2.4.1 Questionnaire

Before the experiment, subjects were asked a number of questions about factors possibly related to performance in the simulator. The questionnaire revealed that only 3 out of 32 subjects were licensed to drive a light motorcycle (< 50cc) and only one of them actually owned one. Only 8 subjects had concrete plans to take driving lessons (within 6 months), whereas 18 subjects did not have any plans yet. Only one person was in possession of a valid theory license.

Most (25) subjects rarely played computer games. Among the games played, car racing was the most popular although none of the subjects used a steering wheel while playing; games were controlled with keyboard and mouse predominantly.

No more than 12 subjects were able to name all eight traffic signs correctly. Nine persons missed only one traffic sign (most often the first sign in appendix C: ‘approaching intersection, give right of way’). At the other extreme of this scale 5 subjects scored only 50% (4 out of 8 signs correct).

6.2.4.2 Practice

For each phase, the number of instructions was compared between groups.
Only in phase 1 the difference in amount of instruction was significant ($t = 2.45$, $p = 0.02$). In the ‘augmented’ condition, subjects received more instructions than in the ‘verbal’ condition.

Although the t-tests for the other three phases did not yield significant results *Figure 15* shows that the percentage of instruction[27] in the verbal condition never exceeded the amount of instruction in the augmented condition.
The changes in the number of instructions over phases generally showed a similar pattern for both conditions. The number of instructions decreased when traffic was added (phases 2, and 4) compared to the previous phase without traffic. An increase in instructions was observed in the third phase when gear changing was added. See also *Figure 15*. Closer inspection of the data revealed that the increase was caused by instructions with regard to speed and lateral position.

6.2.4.3 Test

No statistical analysis was done to compare the number of subjects that passed or failed in each phase. The reason for this is that only three (out of 16) subjects in the augmented instruction condition completed phase 3 and 4 successfully. In the verbal instruction condition, 8 subjects made it through the third phase and 6 of them also performed satisfactorily in the final phase of the experiment. *Figure 16* shows the percentage of subjects who passed each phase.

Figure 15. *The relative amount of instruction received by the subjects in the different conditions.*

[27] Because of differences in the number of trials per phase and per condition, presentation of absolute data would not be illustrative.

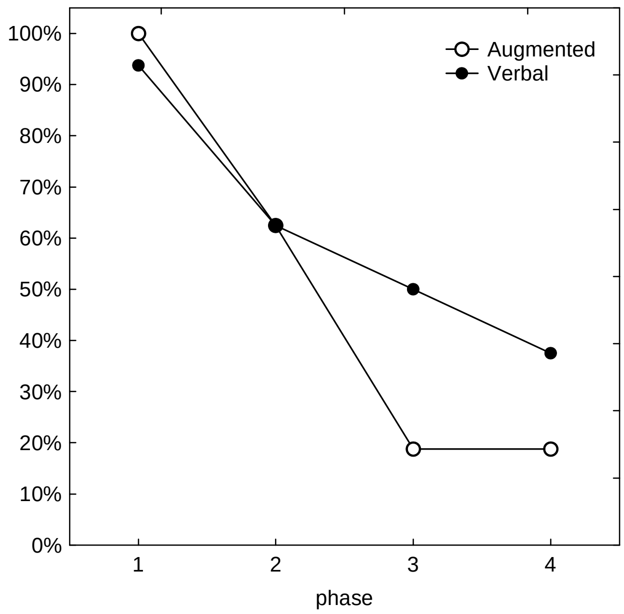


Figure 16. *Percentage of subjects in both conditions that passed the experimental phases.*

The performance measures (dependent variables) were investigated for differences between conditions within phases by means of analysis of variance. From the first MANOVA (2-condition × 2-phase × 3-road-type) it can be seen that the three main effects are significant. Since we were not interested in the differences between the phases and road types, only the main effect of condition is described here: $F(6,175) = 2.539$, $p = 0.022$). This effect was completely caused by the variability of speed (SD speed) which was significantly higher for the augmented instruction group $F(1,180) = 12.532$, $p < 0.001$.

The only significant interaction between phase and road type is not discussed further because it did not involve the factor condition. This interaction is to be expected because driving behavior in curves and on straight roads (road type) is affected differently by the changing requirements of the subsequent phases. For example, trainees had to drive faster on the straight roads as soon as they learnt how to shift gears. In curves, however, they always had to drive slowly.

The MANOVAs for phase three and four respectively only gave a significant difference for factor road type. Because of the inherent differences between the road types (straight roads and curves) this result will not be discussed further either.

6.2.5 Discussion and Conclusions

The instructor was very positive about most of the augmented instructions because they seemed to work intuitively and conveyed meaning instantaneously whereas verbal instructions took more time to process. Only in a few cases (mostly concerning the aurally presented augmented instructions) she would have preferred a verbal instruction because the augmented variant was somewhat contrived. Strikingly, these were instructions for which the timing was less important than the content (i.e., subjects did not violate traffic rules or were not involved in a dangerous situation otherwise) for example the instructions for shifting gears up or down, or driving too slow. It appears that as soon as a trainee has to 'translate' a cue to find out what it means, a verbal cue is probably easier to understand than an augmented one.

Most instructions that were given referred to the lateral position on the road. The overall percentage of instruction showed a decrease over the experiment. We have to be cautious to attribute this to training because of task differences between phases. In favor of a training effect is that the basic control aspects stayed comparable and the level of difficulty increased rather than decreased for each subsequent phase. An alternative explanation that the decrease originated from ‘instructor fatigue’ is implausible because the decrease is non-monotonic: the amount of instruction showed a slight increase in phase 3.

Augmented instructions were delivered more often than verbal instructions throughout the experiment. However, this difference was significant only in the first phase of the experiment. Two plausible explanations can be put forward for this difference between conditions.

Augmented instruction is less ‘intrusive’ than verbal instruction, e.g.: showing a traffic sign with the speed limit is more friendly than ‘saying’ ‘you’re driving too fast’ especially when the instruction is given five times in a row. Therefore, the instructor felt more comfortable giving the augmented instructions. However, some of the augmented instructions were actually used *less* often than their verbal counterparts. It turned out that these were conveying a more complex message such as 'gear up'. In these cases speech may be more effective to pass the message.

The second explanation is that augmented instruction simply did not work as good as verbal instruction. Support for this explanation comes from the fact that more instruction and feedback was provided in the augmented condition as compared to the verbal condition while at the same time there were less trainees in the augmented condition who completed all phases of the experiment. Unfortunately, we could neither confirm nor reject this explanation because of insufficient material to compare both groups.

Post hoc, after discussion with the instructor, we came up with a third explanation. It seemed that a relatively large part of the subjects assigned to the ‘augmented condition’ were already less skilled than the subjects in the ‘verbal condition’ prior to the experiment. Although the results of the questionnaire gave no reason to suspect differences in (driving) ability prior to the experiment, considerable differences in proficiency were observed during driving in the simulator providing indirect support for this claim. As far as could be determined, however, differences were not related to sex, computer gaming experience, exposure to traffic in everyday life, knowledge of traffic rules, or age.

Even though differences in amount of instruction (practice trials) between groups do not necessarily reflect differences in subsequent performance (test trials) in this case the data provide at least some evidence that more instruction is related to worse performance: in the first phase of the experiment, the subjects in the augmented instruction condition received more instructions compared to the subjects in the verbal condition. In the subsequent test trial their performance was also worse. This was reflected mainly in the variability of their speed control. In the first phase, the 'verbal' group was better able to maintain a constant speed. Although this difference was not found to be significant in any of the other phases of the experiment, still it was observed that only three subjects in the ‘augmented’ condition were able to reach the final phase of the experiment against eight in the ‘verbal’ condition.

Extrapolating from these findings, it may be hypothesized that subjects that have to invest much effort in one aspect of driving will have less (attention) capacity to invest in other aspects (keeping a steady speed in this case). As a result, the performance on other task aspects or even over all task performance may deteriorate. This line of reasoning is also in line with the idea that the subjects in the augmented instruction condition were already less able or less talented to learn to drive a car than the subjects in the verbal instruction condition.

Based on the available data it cannot be determined with any certainty whether these results should be attributed to training factors (i.e. the augmented instructions were less effective) or a priori differences between subjects. Both explanations can be defended.

6.3 Experiment 3: Interaction of mode with aptitude

Because the data of Experiment 2 did not lead to unambiguous conclusions with regard to the differences between the two types of instruction, a second experiment was conducted with the same experimental set up. In order to reduce the possibility of disturbing effects of a priori skill levels a preliminary test was introduced to assign the subjects to two groups (one high aptitude group and a low aptitude group). The two instruction conditions of the first experiment were also maintained resulting in a four-group design (2 aptitude x 2 instruction).

Table 10. *Experimental groups (instructional condition x aptitude)*

Experimental groups		**Instruction condition**	
		Augmented	Verbal
Aptitude (estimated from test drive)	Low	Low-augm.	Low-verbal
	High	High-augm.	High-verbal

To rule out the effect of a disturbing variable, subjects should be matched on that variable. However, a first requisite is the availability of a preliminary test that can give a reliable indication of the subjects' scores on the matching variable. In the present experiment, the relevant question is 'how can we match subjects on their aptitude for driving the simulator?' Many standard cognitive and psychomotor tests have been proposed and used to give an indication of driving performance (Ball, Owsley, Sloane, Roenker, & Bruni 1993; Heikkilä, 2000). A few examples are 'visual acuity, contrast sensitivity, eye health, visual memory, personality questionnaires, (choice) reaction time, and information processing tests. Correlation with driving performance (investigating crashing behavior, Ball et al., or looking at faults and offences, Heikkilä) is generally low (indications) although studies with specific groups of drivers have been able to yield higher correlations with driving performance. Considering it is difficult to generalize from findings referring to people with neurological deficits or patients suffering from Parkinson's disease how to predict the performance of people who never have driven before. The usability of this measure therefore is questionable. Furthermore, it is known that there are large differences between inexperienced and experienced drivers with regard to their ability to select the appropriate cues and to interpret them correctly (Summala, Lamble, & Laakso, 1998).

Korteling (1994) already recognized this problem and stated that there is no better way to predict people's performance on a task than by means of the task itself. This is exactly what we decided to do. The subjects were assigned to the conditions (verbal instruction or augmented cues instruction) based on their performance on a short preliminary test in the simulator. An experienced driving instructor judged their performance.

6.3.1 Method

6.3.1.1 Instrumentation

The simulator and database used in this experiment were exactly the same as in the previous experiment. No changes were applied. *Figure 17* clearly shows the semi-circular configuration of five displays and the use of LCD-screens for mirrors. *Figure 18* gives a more detailed view on the LOCS dashboard: an LCD-screen, which shows a speedometer, odometer, rev meter, gear indicator, and indicator lights. A schematic view of the configuration can be found in *Figure 12* in this chapter.

Figure 17. *The TNO-HF Low-cost simulator (LOCS).*

Figure 18. *Detailed view of the LCD-screen with dashboard information*

6.3.1.2 Subjects

Twenty-eight subjects participated in this experiment (11 male, 17 female subjects). These were recruited from the "TNO subject data bank" just like in the previous experiments. Based on their preliminary test results, subjects were assigned to the experimental groups. The procedure that was used for this is explained in the next paragraph. Eventually seven subjects were assigned to each group. In the first place, care was taken that the number of high and low skill subjects was equal for both conditions. With regard to the ratio of male / female subjects in each group it can be said that this was correctly represented in both low aptitude groups. Female subjects, however, were slightly over-represented in the high aptitude - augmented instruction group.

The average age of the subjects was 21.1 years with a standard deviation of 2.7 years (youngest 18, oldest 30). None of the subjects had received driving instruction in a car prior to the experiment.

6.3.1.3 Conditions

Just as in the previous experiment, there were two types of instruction, both activated by keystrokes: the verbal instruction condition in which recorded voice-commands were played as instructions, and the augmented instruction condition in which these commands were translated into a non verbal-instruction. Additionally, a division was made by means of performance on a preliminary simulator-driving test. In line with the results of this test, subjects were assigned to a low aptitude group or a high aptitude group.

In order to match the subjects on their aptitude to drive a simulator it was decided to have the subjects drive two short routes prior to the experiment (approx. 300 meter with 3 curves / intersections). During those test trials no experimental instruction was given. Afterwards the instructor gave an estimation of the ease with which a subject would learn the task. Subsequently the subjects were assigned to one of the conditions (verbal instruction or augmented cues instruction). Subjects did not know what conditions the experiment comprised.

In both conditions subjects received the same four phases of a practice run followed by a test. During practice the instructor provided help (different depending on instructional condition). During the test no instruction was given (so that the tests were the same for both groups). After the test, the (subjective) instructor judgment was used to decide whether the trainee should continue to the next phase or repeat the same phase.

Each phase introduced some additional difficulties: Initially, the trainees were supposed to drive (at low speed) without shifting gears. No other cars drove around. In phase 2 other traffic was present. Traffic was removed again in phase three but gear shifting was introduced here. Finally in phase 4 trainees had to deal with other traffic while shifting gears.

The four groups resulting from this manipulation will be referred to as AI-low or AI-high and VI-low or VI-high (where AI stands for augmented instruction and VI for verbal instruction).

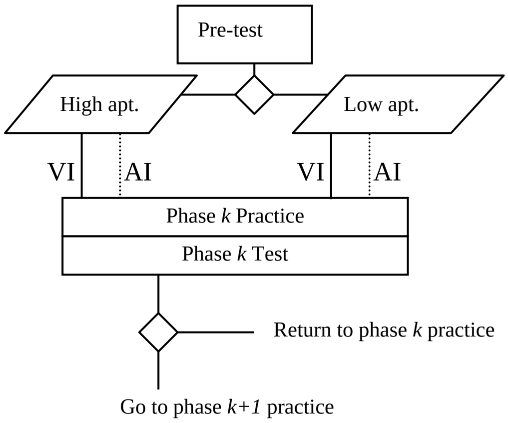


Figure 19*: Experimental design. Diamonds represent a decision of the instructor. VI means 'Verbal Instruction', AI means 'Augmented Instruction'. After the pre-test, all subjects start in phase 1.*

6.3.2 Hypotheses

The main question in this experiment was if the two instructional conditions would have different results. Augmented instruction was expected to be better than verbal instruction particularly when care was taken to rule out aptitude as a disturbing variable. It was hypothesized that differences in skill between groups of trainees would be reflected in the amount of instruction they received during practice, and in their performance during the test trials. Furthermore, we expected subjects that were classified as 'high-aptitude', would perform equally well regardless of the instructional condition they were assigned to. Low aptitude subjects however were expected to profit more from augmented instruction compared to verbal instruction.

6.3.3 Analyses

6.3.3.1 Practice

Within each phase, a comparison of the (relative) amount of instruction between the experimental groups (aptitude - treatment) was made with an ANOVA. This was done because the absolute amount of instruction was not comparable due to differences in the time per phase between subjects.

Just as in experiment 2 the differences in amount of instruction over the experiment (i.e.) between the phases were not tested statistically because not every instruction (or feedback message) could be given in each phase. Instead, a qualitative interpretation of these differences is provided based on visual inspection of the data.

6.3.3.2 Test

To evaluate the performance measures, in phase 1 and 2 a MANOVA was used. An ANOVA was conducted in phase 3. Because none of the subjects in the category 'augmented / low' participated in the fourth phase, no analyses were done for this phase.

6.3.4 Results

6.3.4.1 Questionnaire

Just like in the previous experiment, a short questionnaire was administered prior to the experiment. Of the 28 subjects, only one was licensed to drive a small motorcycle (<50 cc) he also owned one. None of the subjects had a theory license for car driving although 3 of them planned to take driving lessons within the next 1 to 6 months. Half of the subjects did not have any of such plans at all.

Only 13 subjects played computer games (four of them regularly). Car racing and '3D shoot 'em ups' were both mentioned five times. None of the subjects used a steering wheel while playing.

The number of subjects that was able to name all eight traffic-signs correctly was 11. Twelve persons failed to name two or more signs. (The maximum number of missed or incorrectly named signs was four).

6.3.4.2 Practice

Only in the first phase the ANOVA yielded significant results, $F(3,24) = 3.379$, $p = 0.03$. A post hoc test (Tukey's HSD) revealed that the AI low group received significantly more instructions than the AI high group.

The differences between the AI-low and the other groups failed to reach significant values although they were considerably far apart. This may be caused by the fact that Tukey's HSD is a rather conservative post-hoc test in that it offers a high amount of protection against the increased alpha error rate due to multiple post hoc comparisons. (For example, with the Newman-Keuls or Fisher LSD post-hoc test all three groups differ significantly from AI-low.)

The ANOVAs for the other phases[28] are not significant. Hence, no further tests were conducted. The data can be seen in *Table 11* showing both the absolute number of instructions as well as the amount of instructions as a percentage of time. When looking at the first row of data in the table (AI Low) it stands out that the relative amount of instruction does not change much (19,56 % - 23,53 %). The absolute number of instructions however, shows a dramatical decrease because of the small number of subjects making it to the third phase. For this reason the percentages were used in the analysis (see also *Figure 20*).

[28] In the ANOVA of the fourth phase a comparison was made between three groups only because none of the AI Low subjects reached the fourth phase of the experiment.

Table 11. *Absolute amount of instructions per phase for each group (percentages between brackets)*

Phase / Group	1	2	3	4
Low-augmented	166 (21,59%)	116 (19,56%)	17 (23,53%)	- (-%)
High-augmented	61 (12,72%)	86 (15,13%)	108 (33,19%)	24 (13,07%)
Low-verbal	81 (13,03%)	72 (13,64%)	103 (42,08%)	11 (23,83%)
High-verbal	52 (13,19%)	57 (10,39%)	124 (30,14%)	28 (17,89%)

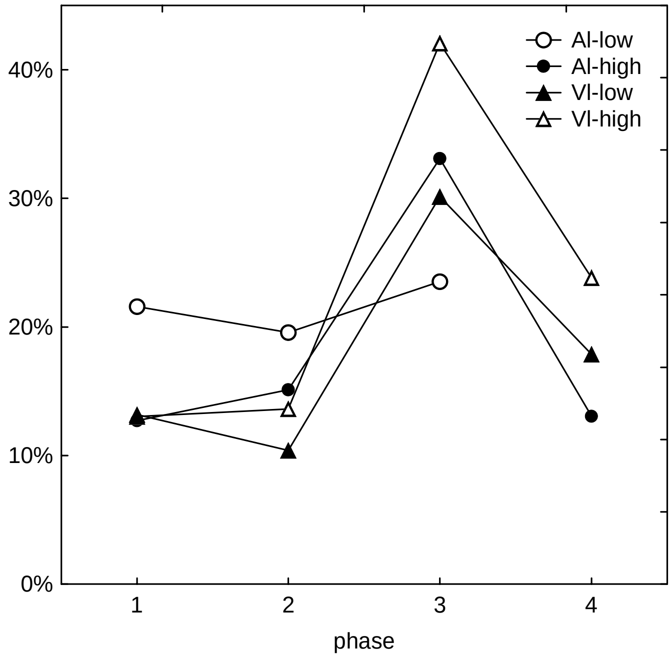


Figure 20. *Relative amount of instruction*

6.3.4.3 Test

An ANOVA on the overall performance of trainees only showed a significant difference between high and low aptitude trainees, $F(6,19) = 3.936$, $p = 0.01$. That is, the high aptitude trainees performed better than the low aptitude trainees did. Remember that the difference between high and low was based on two short runs in the simulator. The difference between the two instruction conditions failed to reach significance. The objective data now clearly confirm the initial judgment of the instructor and indicate that she was not biased when assigning trainees to the categories or judging their performance. This is illustrated in *Table 12* which shows the number of subjects per aptitude-treatment condition and their highest level of performance based on the instructor judgment.

Table 12. *Aptitude treatment and performance*

	Level 1	Level 2	Level 3	Level 4
Low-augmented	0	6	1	0
High-augmented	0	1	3	3
Low-verbal	0	3	3	1
High-verbal	0	0	4	3

With regard to the instruction conditions no significant differences were found on the performance measures in phases 1, 2, and 3. Phase 4 did not have enough data to compare the two instruction groups. The conditions were compared per road type on each of the performance measures: speed, lateral position, and turning rate (for each of these a 5 sec average and standard deviation)

6.3.5 Discussion and Conclusions

After the second experiment had been analyzed, it was hypothesized that a confounding variable (aptitude) possibly obscured the (positive) effects of the augmented instruction condition. This (third) experiment was set up to control for differences between subjects as much as possible by matching subjects on their aptitude for learning to drive.

For this means, the instructor was asked to give an estimate of each subject's aptitude based on their performance during a short driving test in the simulator[29]. Within such a short period of observation, the instructor was able to assign each subject to either the high- or the low-aptitude group. Despite the subjective nature of this procedure, the experimental data (from the following sessions) clearly confirmed the instructor judgment. It proved impossible to distinguish between subject aptitude based on (objective) demographic data, or knowledge of traffic rules as tested in a questionnaire that was also administered prior to the experiment. Contradictory as it may seem, this finding is in line with the results from studies that investigated a number of standard cognitive and psychomotor tests to give an indication of driving performance of neurological patients during the process of rehabilitation (e.g. Ball, Owsley, Sloane, Roenker, & Bruni 1993; Heikkilä, 2000). Test measuring visual acuity, contrast sensitivity, eye health, visual memory, personality questionnaires, (choice) reaction time, and information processing tests fail to show a high correlation with 'fitness for driving'. It seems that the best test for 'fitness for driving' is the driving task itself. The use of a driving simulator would allow for standardization of such a test in a safe environment.

[29] Note that this was done prior to the experiment. To be able to complete the test, the subjects, who did not have driving experience, needed some instruction on how to drive in the simulator. This could be kept very limited and did not seriously affect performance during the experiment.

With regard to the data it was observed the amount of instruction in the first phase was considerably lower in the present experiment than in the previous (second) one. Instead of the gradual decrease in instruction that was observed earlier, now it rather stayed on an even level with a temporary increase in the third phase. It is difficult to come forward with an explanation for this observation.

Just as in the second experiment, the amount of instruction delivered to the augmented group was larger than to the verbal group. However, this difference was only significant in the first phase and it was caused by the low aptitude trainees. The fact that the verbal aptitude groups did not show a difference might be seen as support for the idea that augmented instruction may be easier to provide when necessary. If there is no need to provide additional instruction, it is used just as much as verbal instruction. This latter type of instruction seems to lack this flexibility. For low aptitude trainees then, it can be a solution to provide non-verbal (augmented) instructions because they seem to be friendlier.

Indeed, none of the subjects in the augmented group complained about the large amount of instruction they received. This may indicate that augmented instructions are less intrusive than verbal instructions and therefore may be easier to administer (or receive).

In the first of these three experiments it was already seen that an error is not the sole criterion for a human instructor to give instruction or feedback. Only on about one out of six possible occasions, an error resulted in an instructor comment. One reason for this selective provision of feedback would be that the instructor would constantly be talking and the trainee would not have time to process the remarks / or direct attention to the (driving) task that is difficult anyway. Another reason may be that the instructor expects the trainee to learn recover from his own mistakes. Verbal instructions can be experienced as a criticism whereas augmented instructions can be seen as a hint leaving open the way to self recovery by the trainee. Therefore, they may be seen as especially helpful to low-aptitude trainees.

A final observation with regard to the augmented instructions was that the amount of AI for the low-aptitude group was smaller than for the other groups only in the third phase of the experiment. This contradicts the assumed flexibility of AI. After all, low aptitude trainees were expected to receive more instruction than their high aptitude counterparts. As a tentative post-hoc explanation it may be suggested that the instructor gave up on these low aptitude subjects. This seems to be supported by the fact that none of the trainees in this group reached the fourth phase.

No significant differences were found in performance in the test phases (even though there was a difference in the amount of instruction). The only difference that was found was in aptitude. Subjects categorized as low aptitude performed worse than high aptitude subjects. This observation, although of no importance to the experimental conditions, confirms the objectivity of the instructor in judging performance during the experiments.

The interaction between aptitude and condition was never significant during the test phases, which leads to the conclusion that there is no reason to treat the aptitude groups differently.

Although no definite conclusions can be drawn from these data, a trend was visible showing that low aptitude subjects performed slightly better with verbal instructions. This was something not expected as the augmented instructions were designed to be easily processed and therefore be beneficial in particular to those subjects that were having trouble to perform the task in the first place. Apart from that the (subjective) reactions of trainees were very positive with regard to the augmented cues.

Possibly, this surprising trend is a consequence of the differences between the twelve instructions. While some augmented instructions may have worked very good, others may have been difficult to interpret. This may have given the verbal instructions the overall advantage. Although these speculations could not be tested with these data, they may be an interesting topic for future research.

One thing this experiment may have cleared up though is the question that remained open after the second experiment: It was suggested that the augmented instruction in that experiment could have yielded better results were it not for the coincidental assigning of low aptitude subjects to the augmented instruction group. This suggestion received support in the final experiment: After the subjects were matched on aptitude prior to assigning them to one of the experimental conditions, no significant differences were found between the instruction conditions (within the groups based on aptitude). The differences between the aptitude groups were significant, however.

It could be true in general that the relatively small effect sizes of experimental manipulations with regard to instructional strategies are easily obscured by inter-individual differences. This is a problem that cannot be solved easily. Experiments of this type could never be done in a 'within subject' design because of the transfer between sessions. For this reason, it might be more fruitful for future research to focus on a more detailed level of the driving task. For example, a researcher could pick out one particular instruction to compare different forms of. A disadvantage of this approach is that it is very laborious. Besides, it would require a task that is more abstract than the current driving task. This would also reduce the validity of the experimental environment and restrict the range of the conclusions.

Chapter 7

Discussion and Conclusions

High-performance tasks are defined in this thesis as complex, time-critical tasks where the operator is in the primary control loop of the system (chapter 1). In the domain of high-performance tasks there is much to be found out about training and instruction by means of simulation. Although many concepts have been thought of, few of them have been empirically tested. Even though there is a strong need for a comprehensive theoretical framework that helps to predict the duration of the learning process as a function of skill level, task characteristics, instructions, and other mediating factors, current knowledge of training does not seem to be able to provide an answer as was seen in the reviewed literature. In general it was concluded that most approaches lack the power to predict why, how and when changes in skills as a result of learning will take place.

The need for such a model can also clearly be seen in the field of training simulation. As far as training simulation is considered, too often the focus is on fidelity: a great deal of time and effort is spent to make sure the simulator approaches reality as much as the current state of technology allows whereas the focus should be on creating an optimal training environment. Such a fidelity approach often leads to attempts to validate a model without reference to the behavior of the trainee. This can be compared to determining the value of an educational program by counting the number of spelling errors in the instruction book.

Not surprisingly then, it was decided to look further than physical fidelity of the simulator in this dissertation. The suggestion from literature that deviation from reality might increase the efficiency of training simulators was taken as a starting point. In addition to this, an approach was chosen that would be empirically testable. Therefore we did not take a global overview on the curriculum or lesson but focused on the separate events that make up a lesson. At this level of detail instruction is summarized by the sequence Briefing - Tutoring - Debriefing in which the instructor is concerned with the provision of support before, during, and after the execution of a single training activity respectively. Tutoring can be said to be the aspect of training that is the least predictable. It comprises the reactions of the instructor concerning the specific trainee behavior ('on-line') and therefore, it is the most dynamic and time-critical aspect of instruction at this level.

One concept that lends itself particularly well for use in experimental setting is augmented cueing. Augmentation is a way to help the trainee to perceive what is relevant for correct performance by increasing the salience of certain aspects in the (virtual) environment. The available literature suggests that augmentation is a powerful concept at the level of tutoring. Although the presented evidence is not really unequivocal, it also has a strong intuitive appeal. It led us to formulate the following research questions:

- How should tutoring in a simulator take place to increase effectiveness and efficiency of the training process? In particular, is it better to stay to reality as close as possible or can it be beneficial to deviate from reality during training.

Because this question is too general to answer, three (sub) questions have been derived from it:

- How do instructions in a high-performance task relate to the learning process?
- How is the efficiency of training in a simulator affected by two different approaches to tutoring: one 'traditional' method based on verbal instruction, and one 'experimental' method based on augmented cueing and feedback?
- Tutoring is a process that revolves around interaction between tutor and trainee. To what extent then do trainee characteristics determine the efficiency of the tutoring process?

Each of these questions was addressed in an experiment using a low cost driving simulator.

The first experiment was explorative in the way that it did not compare the effects of different experimental manipulations. Instead it comprised a registration and analysis of the instruction process in a driving simulator. As such it has given an indication with regard to the nature of the instructions that are used during driver training and the frequency with which they occurred. It turned out that 70% of all instructions were categorized in one of fourteen categories (leaving 24 other categories nearly useless).

Although it was possible to predict instructor judgment (at the end of a test trial) with a fair degree of accuracy (70 - 100%) using six performance measures, the attempt to predict the instructions themselves (during practice) was only partly successful. The percentage of correctly classified instructions varied between 30 and 40 % (with one exception of 65%) in the different phases of the experiment. Although this was significantly higher than chance level it must be concluded that this was not sufficient for accurate classification of samples in the appropriate instruction categories.

The main cause for this could be found in the fact that there simply was no one-to-one mapping of instructions to errors. Apparently to prevent excessive workload, human instructors leave many mistakes without consequence. Also, it was concluded that the instructor for a large part relied on the self correcting abilities of the trainees. It became apparent that the relatively low level of instruction did not imply that the trainees’ mistakes were not noted. The fact that the instructor judgment at the end of the test trials could be predicted with high accuracy indicated that the overall judgment was constructed around a number of observations of performance. In short, it was observed that an instructor tries to find an optimum between consistency of instruction and workload by including subjective factors such as training history and assumptions about trainee skills when deciding to give instruction or not.

Coming back to the first research question it is clear that the trainees have learned something during this experiment in a relatively short period of time. By the end of the experiment, most of them were able to control the simulated car and drive through the database, on curves and on straight roads, at different speeds, in different gears, and interacting with other traffic. Whether training has been *efficient* could not be determined. Nevertheless, two important findings emerge from the results: the lack of consistency between performance and instruction, and the observation that only a small set of instructions was applied to the majority of errors.

In the second experiment an attempt was made to compare the efficiency and effectiveness of two different modes of instruction: the verbal instructions, which had proven their effectiveness in the first experiment, and a set of augmented cues. The instructor was very positive about the augmented instructions because they seemed to work intuitively and conveyed meaning instantaneously whereas verbal instructions took more time to process. Augmented instructions were delivered more often than verbal instructions throughout the experiment. However, this difference was significant only in the first phase of the experiment. The evidence indicated that this was partly caused by the fact that the augmented instructions were experienced as less intrusive and thus were provided more easily. Another explanation was that, in spite of the attempt to create comparable experimental groups, a-priori skill differences between the two groups were responsible for the increased amount of instruction with regard to augmented cues. As far as could be determined, however, differences were not related to sex, computer gaming experience, exposure to traffic in everyday life, knowledge of traffic rules, or age. Nevertheless, the number of trainees successfully completing the experiment was clearly highest in the verbal condition. With this explanation in mind, the second research question could not be answered satisfactorily. At first sight, the suggested benefit of augmentation was not supported by the data. But an alternative explanation could not be ruled out.

The third experiment was set up to control for differences between subjects as much as possible by matching subjects on their aptitude for learning to drive.
Again, the amount of instruction delivered to the augmented group was larger than to the verbal group. However, this difference was only significant in the first phase and strongly related to the low aptitude group. The data provide further evidence for the suggestion that augmented instruction is easier to provide than verbal instruction. If there is no need to provide additional instruction, it is used just as much as verbal instruction. This latter type of instruction seems to lack this flexibility. For low aptitude trainees then, it can be a solution to provide non-verbal (augmented) instructions because they seem to be friendlier. Nevertheless, the low aptitude trainees in the verbal condition did not perform worse than their counterparts receiving augmented cues. This is attributed to the differences between the different instructions. While some augmented instructions may have worked very well, others may have been difficult to interpret. This may have given the verbal instructions the overall advantage. Although these speculations could not be tested with these data, they may be an interesting topic for future research.

One thing this experiment may have cleared up though is the question that remained open after the second experiment: It was suggested that the disappointing results of augmented instruction in that experiment could have been due to the coincidental assigning of low aptitude subjects to the augmented instruction group. This suggestion seems to be confirmed in the final experiment: After the subjects were matched on aptitude prior to assigning them to one of the experimental conditions, no significant differences were found between the instruction conditions (within the groups based on aptitude). The differences between the aptitude groups were highly significant however.

When we try to connect these results to the research questions it stands out that the relation between instructions and learning process is a complex one. It became clear that no one-to-one relationship existed between errors that were observed during task performance, and the support as provided by the instructor. In fact, the amount of instruction was clearly lower than would be expected on basis of performance alone. There are two aspects to this observation: First, the instructor might rely for a part on the self correcting ability of trainees. This is in fact a way to slowly reduce the scaffolding that comprises the tutoring process (Merrill et al., 1992). Second, there is a risk of cognitive overload (both for the instructor as for the trainee) when increasing the amount of instruction. Our findings indicate that augmented cues can be provided with a higher intensity than verbal messages. According to Tabbers (2002), however, the modality effect would predict that spoken instruction helps to prevent cognitive overload when the task contains much visual information. As these findings were accumulated from research in the field of desk-top simulation (e.g., multi media learning about geometry) it has to be awaited in how far they can be generalized to high-performance tasks. Interesting in this light, would be to investigate the difference in efficiency between printed messages (as they are often used in multi-media learning) and the augmented cues used in the current research.

Taking a different perspective one also might consider to search for rules when not to disturb the driver as to keep workload within acceptable margins. Research on in-car support systems for example (Verwey, 1990; 1991) has shown that under certain conditions (interaction with other traffic) is detrimental to task performance. In a training environment, similar guidelines could be derived to create intelligent support systems that may, for example, postpone feedback messages or present them in a modified (less attention demanding) form.

- How is the efficiency of training in a simulator affected by two different approaches to tutoring: one 'traditional' method based on verbal instruction, and one 'experimental' method based on augmented cueing and feedback?

Even though it was difficult to derive specific guidelines about the instructional interventions that should be used at the tutoring level from the theoretical models that were reviewed, Romiszowski (1981; 1999) states that feedback (knowledge of results and knowledge of performance) is a very important mechanism for fine-tuning of reproductive aspects in skill acquisition. In this light an interesting line for future research would be to investigate if the tutoring differs between productive and reproductive (or non-recurrent and recurrent) elements in a high-performance task. As the present research suggests that both augmented cues and verbal instructions may be optimal under specific conditions, some form of blended learning, combining multi-media and face-to-face instruction, might unite the positive aspects of both approaches.

- Tutoring is a process that revolves around interaction between tutor and trainee. To what extent then do trainee characteristics determine the efficiency of the tutoring process?

The role of trainee characteristics may be larger than suspected initially as we found that aptitude obscured the effectiveness of augmented cues. How exactly this works is not yet clear. That aptitude is responsible for a large part of inter-individual differences has been demonstrated convincingly. The reviewed models, however, do not account for this factor.

What is needed therefore is a stronger link of instructional design models with guidelines derived from practice. Eventually several research approaches should converge to establish such a link on all levels of instruction (from event to lesson to curriculum) so that a comprehensive model of high-performance task training can be developed.

Overall, it can be stated that it is very difficult to draw conclusions about the tutoring process. Augmented cues are not per se good or bad. It could be true in general that the relatively small effect sizes of experimental manipulations with regard to instructional strategies are easily obscured by inter-individual differences. This is a problem that cannot be solved easily. Experiments of this type could never be done in a 'within subject' design because of the transfer between sessions. For this reason, it might be more fruitful for future research to focus on an even more detailed level of the driving task. For example, a researcher could pick out one particular instruction to compare different forms of. Especially for continuous tasks such as steering and speed control, continuous feedback mechanisms such as augmentation might be beneficial because there exists a compatibility between the instruction and the appropriate response (see also Wickens, 1992). A disadvantage of this approach is that it is very laborious. Besides, it would require a task that is more abstract than the current driving task. This would also reduce the validity of the experimental environment and restrict the range of the conclusions.

References

Adams, J. A. (1971). A Closed-Loop Theory of Motor Learning. *Journal of Motor Behavior*, 3: 111-149.

Alessi, S. M. and Trollip, S. R. (2001). *Multimedia for learning: Methods and development*. Needham Heights, MA: Allyn & Bacon.

Alessi, S. M. (1988). Fidelity in the Design of Instructional Simulations. *Journal of Computer-based Instruction, 15*(2): 40-47.

Allen, J. A.; Hays, R. T., and Buffardi, L. C. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. *Human Factors*, *28*(5): 497-509.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4): 369-406.

Anderson, J.R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105*(2): 165-180

Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., and Bruni, J. R. (1993). Visual Attention Problems as a Predictor of Vehicle Crashes in Older Drivers. *Investigative Ophtalmology & Visual Science, 34*(11): 3110-3123.

Boer, J.P.A. (1991). *The use of simulators for Education and Training I: Factors that determine the value of a simulator as a learning tool.* [In Dutch: Het Gebruik van Simulatoren voor Opleiding en Training I: Bepalende factoren voor de Waarde van een Simulator als Leermiddel]. (Report No. IZF-91-A048). Soesterberg, The Netherlands: TNO-Institute for Perception.

Boldovici, J. A. (1987). Measuring transfer in military settings. In: Cormier, S. M. and Hagman, J. D. (Eds.), *Transfer of learning* (pp. 239-260). Orlando, FL: Academic Press, Inc.

Boldovici, J. A. (1992). *Simulator Motion.* (ARI Technical Report 961). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A257 683).

Bosch, K. van den, and Riemersma, J. B. J. (2000). Praktijktraining met simulaties [Practical training using simulations].Handboek Effectief Opleiden 23 (2000), nr.7.4-6: 107-132

Breda, L. van, and Boer, J. P. A. (1988). *Validation Study Driving Simulator Leopard II [In Dutch: Validatiestudie Rijsimulator Leopard II]* . (Report No. Memo IZF 1988-M7). Soesterberg, The Netherlands: TNO-Institute for Perception.

Broadbent, D. E. (1987). Structures and Strategies: Where are we now? *Psychological Bulletin, 106*(3): 410-433.

Brown, J. S. (1989). Towards a new Epistemology for learning. In: Frasson, C. and Gauthiar, J., (Eds.), *Intelligent Tutoring Systems at the Crossroad of AI and Education*. Norwood, NJ: Ablax.

Chapman, P., Underwood, G., and Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. Transportation Research Part F 5: 157-167.

Cook, C. A., O'Shea, A. M., Young, A. L., and Stedmon, A. W. (1999). Helpful feedback? Deriving guidelines for advanced training systems. In: *Proceedings of the Human Factors and Ergonomics Society 43rd annual meeting*: 1133-1137.

Duncan, J.; Williams, P., and Brown, I. (1991). Components of Driving Skill: Experience Does not Mean Expertise. *Ergonomics 34* (7): 919-937.

Emmerik, M. L. van, and Jong, T. de (In preparation). *Review of Simulator Specific Training and Instruction Factors.*

Emmerik, M. L. van, Rooij, J. C. G. M. van, and Jong, T. de (2000). *Exploration of Instruction in a Driving Simulator.* (Report No. TM-00-B008). Soesterberg, The Netherlands: TNO-Human Factors.

Ericsson, K. A., and Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*: 273-305.

Farmer, E. W., Jorna, P. G. A. M., Riemersma, J. B. J., Rooij, J. C. G. M. van, and Moraal, J. (1999). *Handbook of Simulator Based Training*. London: Ashgate.

Fitts, P. M., and Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks Cole.

Fleishman, E. A. and Quaintance, M. K. (1984). *Taxonomies of human performance: the description of human tasks*. Orlando, FA: Academic Press.

Gagné, R. M. (1985). *The conditions of learning and theory of instruction (4th edition)*. Tokyo: Holt-Saunders Japan.

Gagné, R. M., Briggs, L. J., and Wager W.W. (1992). *Principles of instructional design(4th edition)*. Fort Worth: Harcourt, Brace, Jovanovich College Publishers.

Gentner, D. R. (1988). Expertise in typewriting. In: Chi, M. T. H., Glaser, R. and Farr, M. R. (Eds.), *The Nature of Expertise*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Gopher, D., Weil, M., and Siegel, D. (1989). Practice under changing priorities: An approach to training of complex skills. *Acta Psychologica*, *71*: 147-178.

Gordon, S. E.; Babbitt, B. A.; Bell, H. H., and Sorensen, H. B. (1994) Development of a real-time simulation with intelligent tutoring capabilities. In: *Proceedings of the Human Factors and Ergonomics Society 38th annual meeting*; *Nashville, Tennessee*: 1247-1251. Santa Monica, California.

Hart, S. G. (1988), Helicopter Human Factors. In: Wiener, E. L. and Nagel, D. C. *Human Factors in Aviation*. London: Academic Press, Inc.

Hayes, N. A., and Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition, 28*: 249-276.

Hays, R. T., and Singer, M. J. (1989). *Simulation Fidelity in Training System Design*. New York: Springer-Verlag.

Heikkilä, V. M. (2000). Relationship of laboratory and on-road tests for driving-school students and experienced drivers. *Perceptual Motor Skills, 90*: 227-235.

Hogue, J,R., Allen, R.W., Rosenthal, T.J., Andersons, F.G., Pelz, C. and Aponso, B. L. (1999). Applying low-cost virtual environments to simulation-based vehicle operator training. *SimTecT99, The Simulation Technology and Training Conference 1999*, Melbourne, Australia.

Ivancic, K. & Hesketh, B. (2000), Learning from errors in a driving simulation: Effects on driving skill and self-confidence, *Ergonomics, 43:* 1966-1984.

Jonassen, D.H., Tessmer, M., and Hannum, W.H. (1999). *Task Analysis Methods for Instructional Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Jong, T. de, (Ed.), (1997). *Supporting authors in the design of simulation based learning environments*. Joolingen, W. R. van; King, S.; Kuyper, M.; Gellevij, M.; Limbach, R.; Pieters, J.; Horst, M. ter; Meij, J. van der; Wieberdink, B., and Veldhuizen, B. van. (SERVIVE project ET 1020 D 8.1). Enschede, The Netherlands: University of Twente.

de Jong, T., van Joolingen, W.R., Swaak, J., Veermans, K., Limbach, R., King, S., and Gureghian, D. (1998). Self-directed learning in simulation-based discovery environments. *Journal of Computer Assisted Learning*, *14*: 235-246.

Jong, T. de and Joolingen, W. R. van (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2): 179-201.

Joolingen, W. R. van (1993). *Understanding and facilitating discovery learning in computer-based simulation environments.* Doctoral dissertation. Eindhoven, The Netherlands: Eindhoven University of Technology;

Kappé, B., Winsum, W. van, and Wolffelaar, P. van (2002). A cost effective driving simulator. In: *Proceedings of the 7th Driving Simulation Conference (DSC2002)*. Paris, France.

Kappé, B., and Korteling, J. E. (1998). *Low-Cost Simulators 3b: Minimal Functional Requirements for a Driver Training Simulator.* (Report No. TM-98-A058). Soesterberg, The Netherlands: TNO- HFRI.

Korteling, J. E. (1990). *Proposal for a New Performance Evaluation and Feedback System for the Leopard 2 Driving Simulator.* (Report No. IZF 1990 A-40). Soesterberg: TNO Institute for Perception.

Korteling, J. E. (1991). *Ergonomic Design of the Instruction Consoles for the Leopard 2 and YPR-765 Driving Simulators.* (Report No. IZF 1991 A-42). Soesterberg, |The Netherlands: TNO-IZF.

Korteling, J. E. (1994) *Multiple-task Performance and Aging*. Doctoral dissertation, Groningen, The Netherlands: Rijksuniversiteit Groningen.

Korteling, J. E. and Sluimer, R. R. (1999). *A critical review of validation methods for man-in-the-loop simulators.* (Report No. TM-99-A023). Soesterberg, The Netherlands: TNO-Human Factors.

Korteling, J. E., and Randwijk, M. J. v. (1991). *Simulators and Training Grounds for Military Driver Training; Literature Survey and Advice [In Dutch*: *Simulatoren en Verkeersoefenterreinen in de Militaire Rijopleiding; Literatuurstudie en Advies.]* (Report No. IZF 1991 A-11). Soesterberg, The Netherlands: TNO-IZF.

Korteling, J. E., Helsdingen, A., and Baeyer, A. v. (2000). *Handbook of Low-Cost Simulators.* (Technical Report WP5.g, ELSTAR, EUCLID RTP 11.8).

Korteling, J. E.; Bosch, K. van den, and Emmerik, M. L. van (1997). *Low-cost simulators 1a: literature review, analysis of military training, and selection of task domains.* (Report No. TM-97-A035) Soesterberg, The Netherlands: TNO- Human Factors.

Kuiken, M. J. (1996). *Instructional support to drivers: The role of in-vehicle feedback in improving driving performance of qualified motorists*. Doctoral dissertation. Haren, The Netherlands: Verkeerskundig Studiecentrum, Rijksuniversiteit Groningen.

Landsheere, V. de (1989). Taxonomies of Objectives. In Eraut, M. (Ed.), *The International Encyclopedia of Educational Technology*. Oxford, Pergamon Press: 352-362.

Lazonder, A. W., Biemans, H. J. W. and Wopereis, I. G. J. H. (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science 51*(6): 576-581.

Lesgold, A. M.; Rubinson, H.; Feltovich, P.; Glaser, P.; Klopfer, D., and Wang, Y. (1988). Expertise in a Complex Skill: Diagnosing X-ray Pictures. In: Chi, M. T. H.; Glaser, R., and Farr, M., (Eds.), *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lintern, G., and Koonce, J. M. (1992). Visual augmentation and scene detail effects in flight training. *The International Journal of Aviation Psychology, 2*(4): 281-301.

Lintern, G., and Roscoe, S. N. (1982). Visual cue augmentation in contact flight simulation. In: S. N. Roscoe (Ed.), *Aviation psychology*. Ames, Iowa: The Iowa State University Press.

Lintern, G., Roscoe, S. N., Koonce, J. M. and Segal, L. D. (1990). Transfer of Landing Skills in Beginning Flight Training. *Human Factors, 32*(3): 319-327.

Lintern, G., Roscoe, S. N., and Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. *Human Factors, 32*(3): 299-317.

Lintern, G., Sheppard, D. J., Parker, D. L., Yates, K. E., and Nolan, M. D. (1989). Simulator Design and Instructional Features for Air-to-Ground Attack: A Transfer Study. *Human Factors, 31*(1): 87-99.

Lintern, G., Thomley-Yates, K. E., Nelson, B. E., and Roscoe, S. N. (1987). Content, Variety, and Augmentation of Simulated Visual Scenes for Teaching Air-to-Ground Attack. *Human Factors, 29*(1): 45-59.

Mané A., and Donchin, E. (1989). The space fortress game. *Acta Psychologica, 71*: 17-22.

Masthoff, J. F. M. (1997). *An Agent-Based Interactive Instruction System*. Doctoral dissertation. Eindhoven, The Netherlands: Eindhoven University of Technology.

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

Merrill, D. C.; Reiser, B. J.; Ranney, M., and Gregory Trafton, J. (1992) Effective tutoring techniques: a comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences. 2*(3): 277-305.

NADS (2004) The National Advanced Driving Simulator. http://www.nads-sc.uiowa.edu/ : www document.

Naylor, J. C., and Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology, 65*(3): 217-224.

Norman, D. A., and Shallice, T. (1980). *Attention to Action: Willed and Automatic Behaviour* (CHIP Report 99). San Diego, CA: University of California, San Diego.

OECD (1976). *Driver instruction*. Paris: Organisation for Economic Co-operation and Development.

Orlansky, J. (1989). *The military value and cost-effectiveness of training*. AC/243 (Panel 7/RSG.15) D/4. Brussels: NATO.

O'Shea, A. M., Cook, C. A., and Young, A. L. (1999). The use of augmented reality in the provision of training feedback. In: *Proceedings of the 10th International Training and Education Conference (ITEC).* The Hague, The Netherlands.

Padmos P. and Milders, M.V. (1992). Quality Criteria for Simulator Images: A Literature Review. *Human Factors, 34* (6): 727-748.

Patrick, J. (1992). Training: research and practice. London: Academic Press Limited.

Polzella, D. J. (1983). Utility and utilization of advanced instructional features in aircrew training devices: Phase 1 - Tactical Air Command. In: *IEEE National Aerospace and Electronics Conference NAECON 83:* 1108-1115. IEEE.

Poulton (1974). *Tracking skill and manual control*. New York: Academic Press.

Powell, A. (1990). What price the instructor? In: *Proceedings of the 1st International Training and Education Conference (ITEC):* 256-259. Birmingham, UK.

Ramesh, R., and Sylla, C. (1990). A model of instructor training analysis in simulation based flight training. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(5): 1070-1080.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction*. Amsterdam: North-Holland.

Regan, M. A., Deery, H. A. and Triggs, T. J. (1998). Training for attentional control in novice car drivers: a simulator study. In: *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting:* 1452-1456, Chicago, IL.

Romiszowski, A. J. (1981). *Designing Instructional Systems*. New York: Nichols.

Romiszowski, A. J. (1999). The Development of Physical Skills: Instruction in the Psychomotor Domain. In: Reigeluth, Charles M. (Ed.), *Instructional Design Theories and Models: Volume II.* Hillsdale, NJ: Lawrence Erlbaum.

Rooij, J. C. G. M. van, and Korving, F. (1995). Automation of simulator based instruction: Methodology and experimental research. In: *5th ITEC proceedings:* 311-319. The Hague, The Netherlands.

Rooij, J. C. G. M. van (1994). Research on simulator-based training and instruction strategies. In: Brookhuis, K. A., Weikert, C. and Cavonius, C. R. (Eds.), *Proceedings of the HFES Europe Chapter annual meeting:* 73-81. Dortmund, Germany.

Rooij, J. C. G. M. van; Vries, S. C. de; Buitelaar, M.; Ligthart, V.; Brouwer, W.; Roessingh, J. J. M.; Verwey, W. B., and Emmerik, M. L. van (1997). *Functional Simulator Requirements for the Chinook and the Cougar Transport Helicopters.* (Report No. TM-97-A062). Soesterberg, The Netherlands: TNO-HFRI.

Roscoe, S. N., (1991). Simulator qualification: just as phony as it can be. *The International Journal of Aviation Psychology*, 1(4): 335-339.

RVV 1990 / BABW. (1995). *Legislation Traffic Rules and Traffic Signs 1990 [In Dutch: Reglement Verkeersregels en Verkeerstekens 1990 / Besluit Administratieve Bepalingen inzake het Wegverkeer]*. The Hague, The Netherlands: Sdu Publishers.

Schmidt, R. A. (1982) *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL: Human Kinetics.

Schmidt, R. A., and Bjork, R. A. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science, 3*(4): 207-217.

Schneider, W. (1985). Training High-Performance Skills: Fallacies and Guidelines. *Human Factors, 27*: 285-300.

Schneider, W. and Detweiler, M. (1986) Changes in Performance in Workload With Training. In: *Proceedings of the Human Factors Society-30th annual meeting*. Dayton, Ohio, USA: 1128-1132. Santa Monica, CA: HFES.

Schneider, W. and Shiffrin, R. M. (1977) Controlled and Automatic Human Information Processing: I Detection, Search and Attention. *Psychological Review 84*: 1-66.

Shiffrin, R. M. and Schneider, W. (1977). Controlled and Automatic Human Information Processing: II Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review 84*: 127-190.

Sierra. (2000). *Driver's Education '99*. www.sierra.com/sierra/driversed99: www document.

Sluimer, R. R., and Bosch, K. van den (2001). *Simulators in Driving Education: A Field Study into Transfer of Training.* (Report No. TM 01-A-030) Soesterberg, The Netherlands: TNO-Human Factors.

Smith, P. L., and Ragan, T. J. (1999). *Instructional Design*. New York: Wiley and Sons, Inc.

Stanton, N. (1996). Simulators: a review of research and practice. In: Stanton, N. (Ed.), *Human Factors in Nuclear Safety*. London: Taylor & Francis.

Summala, H., Lamble, D. and Laakso, M. (1998). Driving experience and perception of the leaf car's braing when looking at in-car targets. *Accident analysis and prevention*, 30(4): 401-407.

Tabachnick, B. G., and Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper Collins Publishers, Inc.

Tabbers, H. K. (2002). *The Modality of Text in Multimedia Instructions: refining the design guidelines*. Doctoral dissertation. Heerlen: Open University of The Netherlands.

Thoeni, U. A. (1999). Experiences with practical driver training in an advanced driving simulator. In: *Proceedings of the 10th International Training and Education Conference (ITEC)*: 454-458. The Hague, The Netherlands.

Thomson, D. R. (1988). Transfer of training from simulators to operational equipment; Are simulators effective*? Journal of Educational Technology Systems, 17*(3): 213-218.

Underwood, G., Chapman, P., Bowden, K. and Crundall, D. (2002). Visual search while driving: skill and awareness during inspection of the scene. Transportation Research Part F 5: 87-97.

Veltman, J. A. and Korteling, J. E. (1993)*. Evaluation study driving simulators Leopard 2 and YPR 765* [In Dutch: Validatiestudie Rijsimulatoren Leopard 2 en YPR 765.] (Report No. IZF-93-A-43). Soesterberg, The Netherlands: TNO-HFRI.

Verstegen, D. M. L. and Barnard, Y. F. (1999). Supporting the design of simulators from a training point of view. In: *Proceedings of the 20th Interservice / Industry Training Simulation and Education Conference* (pp. 431-441). Arlington, VA, USA: National Training Systems Association.

Verstegen, D. M. L., Barnard, Y. F., and Rooij, J. C. G. M. van (1999). *The Use of Simulators for Training and Instruction.* (Report No. TM-99-B003). Soesterberg, The Netherlands: TNO-HFRI.

Verwey, W. B. (1990) *Skill acquisition in driving and interaction with car-driver interfaces: An exploratory study on determinants of task integration.* (Report No. IZF-1990-C-11) Soesterberg, The Netherlands: TNO-HFRI.

Verwey, W. B. (1991) Adaptive interfaces based on driver resource demands. In: Queinnec, Y. and Daniellou, F. *Designing for everyone: Proceedings of the 11th Congress of the International Ergonomics Association.* London, Taylor & Francis, 1991: 1541-1543

Wachtel, J. A. (1996). Applications of appropriate simulator technology for driver training, licencing and assessment. In: Gale, A. G. et al., (Eds.), *Vision in Vehicles - V.* Amsterdam: Elsevier Science B.V.

Wheatcroft, E. (1973). Simulators for Skill. London: McGraw-Hill.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: Harper Collins Publishers Inc.

Wierda, M. (1993). *Learning effects of driving instruction in a virtual environment* [In Dutch: Leereffecten van een rij-instructie in een virtuele omgeving]. (Report VK 93-14). Haren, The Netherlands: Verkeerskundig Studiecentrum, Rijksuniversiteit Groningen.

Wightman, D. C. and Sistrunk, F. (1987). Part-Task Training Strategies in Simulated Carrier Landing Final-Approach Training. *Human Factors, 29*(3): 245-254.

Winsum, W. van and Korteling, J. E. (1998). *Low-cost simulators 3b: Task analysis for driving simulation*. (Report No. TM-98-A041). Soesterberg, The Netherlands: TNO-Human Factors.

Wolz, U., McKeown, K. R., and Kaiser, G. E. (1990). Automated tutoring in interactive environments: a task-centered approach. *Machine-Mediated Learning 3*, 1: 53-79.

Young, A. L., Stedmon, A. W., and Cook, C. A. (1999). The potential of augmented reality technology for training support systems. In: *People in Control,* Bath, UK.

Zeltzer, D. and Pioch, N. J. (1996). Validation and Verification of Virtual Environment Training Systems. *Virtual Reality Annual International Symposium (VRAIS '96)*, Santa Clara, CA., USA.

Samenvatting (Summary in Dutch)

Er zijn verschillende factoren aan te wijzen die het leren uitvoeren van een taak moeilijk kunnen maken zoals de vereiste snelheid waarmee bepaalde handelingen uitgevoerd moeten worden, de taakstructuur, en omgevingscondities. Wanneer al deze factoren gezamenlijk een rol spelen wordt wel gesproken over 'high-performance taken'. Het besturen van een (gevechts)helikopter wordt in dit kader vaak genoemd als voorbeeld.

Vanwege de bovengenoemde kenmerken zouden high-performance taken goed met behulp van simulators getraind kunnen worden. Simulators bieden namelijk een veilige omgeving en hebben tevens een aantal didactische voordelen. Dit proefschrift richt zich met name op de mogelijkheden om deze didactische voordelen te benutten omdat dat in de praktijk onvoldoende gebeurt. Dit is concreet vertaald in een onderzoeksvraag door gericht te kijken naar één van die voordelen: het gebruik van augmented cues. Specifieke vragen waren:

- Hoe verhouden instructies in een (gesimuleerde) high-performance taak zich tot het leerproces
- Leidt het gebruik van augmented cues en feedback tijdens tutoring tot verschillen in trainingsefficiëntie vergeleken met verbale instructies.
- Is er een interactie tussen tutoringstrategie en aanleg van de lerende.

De autorijtaak wordt in deze dissertatie als voorbeeld van een (in een simulator te trainen) high-performance taak genomen.

In het tweede en derde hoofdstuk worden de gemaakte keuzes toegelicht. Twee begrippen die daarbij centraal staan zijn taken en vaardigheden. Een taak wordt gedefinieerd als een prestatienorm die onafhankelijk is van de uitvoerende. De term vaardigheid verwijst naar de capaciteit die die persoon heeft om de prestatienorm te bereiken. Vaardigheden kunnen wijzigen door training terwijl een taak onveranderlijk is.

Voor het uitvoeren van high-performance taken zijn veel verschillende vaardigheden nodig. Welke dat zijn kan worden achterhaald door middel van taakanalyse: Het splitsen van de complexe taak maakt het mogelijk om vaardigheden aan sub-taken te koppelen en consequenties voor gepaste trainingstrategieën te bepalen. Het ligt immers voor de hand dat deze voor de verschillende soorten vaardigheden eveneens verschillend zijn. Hoewel dit logisch klinkt is er geen gestandaardiseerde manier om vaardigheden aan taken, laat staan aan leerstrategieën, te koppelen. Het lijkt of voor iedere toepassing een aparte categorisatie wordt gemaakt. In veel benaderingen wordt bovendien geen rekening gehouden met de veranderlijkheid van vaardigheden.

In hoofdstuk 2 worden een aantal van dit soort categorisaties besproken. Twee daarvan lijken met name toepasbaar binnen de context van dit proefschrift. Het 4C/ID model van Van Merriënboer (1997) en de 'four stage skills cycle' van Romiszowski (1981; 1999). Beide modellen beschrijven de ontwikkeling van vaardigheden en de mechanismen die hierbij werkzaam zijn.

Van Merriënboer onderscheidt zogenaamde 'recurrent skills', vaardigheden die door het automatiseren van regels kunnen leiden tot snelle en accurate taakuitvoering en 'non-recurrent skills' die via schemaverwerving worden aangeleerd. Gedrag dat op deze vaardigheden leunt kan makkelijk aangepast worden aan veranderende omstandigheden maar is niet altijd snel en accuraat.

Analoog hieraan beschrijft Romiszowski productieve (specifiek afgestemd op een situatie, vgl. non-recurrent skills) en reproductieve vaardigheden (algoritmisch uit te voeren, vgl. recurrent skills). Het model laat zien dat een reactie op een stimulus (taakuitvoering) in eerste instantie langzaam is omdat de hele cyclus bewust moet worden doorgelopen. Productieve vaardigheden moeten deze cyclus altijd doorlopen. Door training zal dit proces gaandeweg sneller worden. Reproductieve vaardigheden echter kunnen door training geautomatiseerd worden. Hierdoor kan de cyclus feitelijk overgeslagen worden zodat de taak uiteindelijk sneller kan worden uitgevoerd dan met productieve vaardigheden.

Om vanuit deze modellen voorspellingen te kunnen doen over het verloop van een leerproces in een simulator is het belangrijk te constateren dat het instructieproces zich op verschillende niveau's afspeelt. Het hoogste abstractieniveau is dat van het curriculum. Hier worden beslissingen genomen die het gehele trainingsprogramma omvatten zoals de samenhang tussen de lessen, en de keuze van het didactische kader. Binnen een les draait het om de samenhang tussen de leermomenten en op het laagste gedetailleerde niveau gaat het om de ondersteuning voor, tijdens, en na de afzonderlijke situaties in een les.

Op elk van deze niveau's is het mogelijk om didactische voordelen van de simulator ten opzichte van het operationele systeem te benoemen. Simulators bieden bijvoorbeeld betere mogelijkheden om: inhoud, structuur en timing van instructie te controleren, de variatie in leermomenten te standaardiseren, de hoeveelheid oefening aan te passen, online prestaties te registreren en te analyseren, en augmented cues (augmentatie) te presenteren. Op dit laatste punt wordt in deze dissertatie dieper ingegaan.

Augmentatie heeft betrekking op het accentueren of toevoegen van (kunstmatig gegenereerde) informatie met als doel bepaalde kritische cues van een taak te accentueren en hierdoor het leerproces te versnellen of de taakuitvoering te vereenvoudigen. In een rijsimulator zou dit bijvoorbeeld kunnen door de kleur van een voertuig dat op botskoers ligt plotseling te veranderen. Augmentatie vindt plaats op het laagste detail niveau van het instructuctieproces, tijdens tutoring (de ondersteuning die tijdens afzonderlijke situaties gegeven wordt), en omdat tutoring een sterk tijd-afhankelijke en dynamische vorm van instructie is sluit het goed aan bij de eveneens tijdkritische high-performance taken.

Opvallend is dat door het gebruik van augmentatie de natuurgetrouwheid (de overeenkomst met de werkelijkheid) van de simulator vermindert. Gesteld wordt dus dat door af te wijken van de werkelijkheid de efficiëntie van het leerproces kan worden bevorderd -een gedachtegang die in de simulatorwereld niet overal geaccepteerd is.

Voor het leren autorijden is simulatie nog niet zo lang in gebruik. De reden hiervoor wordt duidelijk wanneer de hoge eisen die aan de benodigde computers gesteld worden en de complexe omgeving die gevisualiseerd moet worden voor een natuurgetrouwe simulatie afgezet worden tegen de relatief lage kosten van een auto met instructeur. Het mag duidelijk zijn dat een vliegtuig volgens die redenatie veel eerder nut heeft van simulatie. Aan de andere kant worden vele malen meer leerlingen opgeleid tot automobilist dan tot piloot wat impliceert dat een rijsimulator een enorme kosten besparing met zich mee kan brengen als de rijtaak niet volledig, of niet volledig realistisch gesimuleerd hoeft te worden zonder dat dit ten koste gaat van de leerwaarde.

Binnen het Europese ELSTAR project is bij TNO een prototype van een low-cost rijsimulator voor training ontwikkeld (genaamd LOCS). Deze is geschikt voor training van ongeveer 60% van alle rijtaken. De taken die niet in de simulator te trainen zijn, kunnen alle goed getraind worden in een echte auto. De experimenten uit dit proefschrift zijn afgenomen met dit prototype dat bestaat uit een autostoel, drie pedalen, versnellingspook en een snelheidsmeter. De gesimuleerde buitenwereld wordt getoond op vijf 24" monitoren die in een halve cirkel opgesteld staan zodat een beeld van ongeveer 200° getoond wordt. Als zodanig biedt dit platform, met beperkte natuurgetrouwheid, voldoende mogelijkheden om onderzoek te doen naar de didactische factoren die de efficientie van het instructieproces beïnvloeden.

Het eerste experiment dat met de LOCS werd uitgevoerd was voornamelijk exploratief van aard. De instructies die tijdens de uitgevoerde taken in de rijsimulator werden gegeven, werden geregistreerd en geanalyseerd. Deze werden in 38 categorieën geplaatst. Het bleek echter dat 14 categorieën voldeden om 70% van de gegeven instructies te beschrijven. De andere categorieën werden dus relatief weinig gebruikt.

Vervolgens werd op basis van de tijdens het rijden gemeten variabelen geprobeerd een voorspelling te geven van het eindoordeel van de instructrice (over een rit) alsmede de losse instructies die zij gaf. Voor de verschillende ritten was de accuratesse van het statistische model tussen de 70% en de 100%. Wat betreft de afzonderlijke instructies bleek het percentage correct echter een stuk lager (tussen de 30 and 40 % met een uitschieter naar 65%) in de verschillende fasen van het experiment. Hoewel significant hoger dan op basis van kans zou worden verwacht waren deze waarden helaas niet voldoende voor een betrouwbare koppeling van stukjes rijgedrag aan instructiecategorieën.

De belangrijkste oorzaak hiervoor ligt in het feit dat er eenvoudigweg geen één-op-één relatie bestond tussen instructies en fouten. Het lijkt er op dat instructeurs in veel gevallen niet reageren op gemaakte fouten. Gezien het grote aantal instructies dat desondanks gegeven werd is dit echter geen kwestie van luiheid. Waarschijnlijker is dat dit gebeurt om te verkomen dat de instructeur constant instructie en feedback geeft. Dit zou voor zowel leerling als instructeur een onacceptabele hoge werklast opleveren. Daarom vertrouwt de instructeur in een aantal gevallen op het zelf corrigerend vermogen van de leerling. Het bleek duidelijk dat de instructeur de gemaakte fouten wel opmerkte. Dit kwam echter pas in het eindoordeel van een rit tot uiting. Het zou dan ook heel goed kunnen zijn dat er een bovengrens is voor de hoeveelheid instructie die een leerling kan verwerken. Als deze wordt overschreden wordt het leerproces negatief beïnvloed. Subjectieve factoren zoals de voorgeschiedenis van de leerling en veronderstellingen over diens vaardigheden kunnen een rol spelen bij de hoogte van deze bovengrens.

Duidelijk is gebleken dat de leerlingen, tijdens de relatief korte tijd die het experiment duurde, geleerd hebben. Aan het eind van de sessie waren de meesten in staat om de simulator te besturen op routes door de database onder verschillende snelheden, met gebruik van versnellingen en interacterend met ander verkeer. Dit zegt echter weinig over de efficiëntie van de training. Misschien was het op een andere manier beter en sneller gegaan. Desalniettemin komen twee belangerijke bevindingen naar voren: Op basis van het rijgedrag kan niet betrouwbaar voorspeld worden wat voor instructie gegeven zal worden. Er is dus geen sterke samenhang tussen rijgedrag en instructie. Toch werd slechts een beperkte set instructieboodschappen gebruikt om het grootste deel van de fouten af te handelen.

Het tweede experiment betrof een vergelijking van de effectiviteit en efficiëntie van twee verschillende soorten instructie. Hiertoe werden twaalf instructieboodschappen uit het eerste experiment geselecteerd. Van elke boodschap werden twee varianten gemaakt. Een spraakboodschap (deze hadden in het eerste experiment al bewezen effectief te zijn), en een variant die op het principe van augmentatie gebaseerd was. Beide varianten werden aangeboden via de functietoetsen van een computer keyboard (F1 - F12). Afhankelijk van de conditie klonk dan een vooraf opgenomen spraakboodschap of werd een augmented cue gepresenteerd.

De instructrice was erg positief over het gebruik van de augmentatie. Het was intuïtief in het gebruik en kon direct een betekenis overbrengen wanneer verbale instructie meer tijd kostte. Tegelijkertijd werden de augmented cues tijdens het experiment vaker gegeven dan de verbale instructies alhoewel het verschil alleen in de eerste fase significant was. Dit kan op verschillende wijzen verklaard worden. Mogelijk werden de augmented cues ervaren als minder belastend voor en door de leerling zodat ze ook eerder gegeven konden worden dan de verbale instructies. Een negatieve verklaring is dat de augmented cues gewoon minder effectief waren dan de verbale instructies. Het feit dat een aantal proefpersonen uit de augmentatie conditie niet tot de laatste fase van het experiment kwam wijst enigzins in deze richting. Het is echter ook mogelijk dat er - ondanks de pogingen om de groepen zoveel mogelijk gelijk te houden - al voorafgaand aan het experiment verschillen in aanleg tussen de groepen waren die het verschil in de hoeveelheid instructie kunnen verklaren. Voorzover na te gaan hingen deze verschillen niet samen met sexe, computerspel ervaring, blootstelling aan verkeer in het dagelijks leven, kennis van de verkeersregels of leeftijd. Bij alle proefpersonen was vooraf een vragenlijst afgenomen om informatie over deze factoren te vergaren.

Er was dus een derde experiment nodig om hier meer duidelijkheid over te verschaffen. Behalve de vragenlijst die ook hier werd afgenomen werd ook nog een korte rijtest afgenomen voor het eigenlijke experiment. Op basis van de prestatie hierop gaf de instructeur aan of de proefpersonen veel of weinig aanleg hadden. De combinatie van aanleg met instructievariant leverde vier condities op.

Weer bleek in de resultaten dat de augmentatie groep meer instructie ontvangen had. Het verschil werd echter in grote mate veroorzaakt door de proefpersonen met weinig aanleg en het was alleen in de eerste fase significant. Dit geeft ondersteuning voor de eerder geopperde verklaring dat augmented instructies makkelijker te geven zijn dan verbale instructies. In de verbale condities werd de hoeveelheid instructie namelijk niet aangepast voor proefpersonen met weinig aanleg terwijl dat in de augmented conditie wel gebeurde. Voor de proefpersonen met veel aanleg was de hoeveelheid instructie ongeacht de variant gelijk. Het zou dus voor leerlingen die moeite hebben met een taak goed kunnen zijn om augmented instructie aan te bieden. Hiervan kan meer aangeboden worden indien een leerling moeite heeft met de taak zonder dat deze overbelast wordt. Desalniettemin werden er geen significante prestatieverschillen gevonden tussen de beide groepen met weinig aanleg. Een plausibele verklaring die echter in toekomstig onderzoek getest zal moeten worden is dat sommige boodschappen zich beter lenen voor augmentatie terwijl andere beter verbaal gepresenteerd kunnen worden. Dit heeft mogelijk te maken met de hoeveelheid informatie die in de boodschap zit en de complexiteit ervan.

Wat in ieder geval duidelijk naar voren kwam uit dit laatste experiment is dat de verschillen tussen de groepen in het vorige experiment voor een belangrijk deel veroorzaakt werden doordat de proefpersonen in de augmented conditie gemiddeld een lagere aanleg hadden dan de proefpersonen in de verbale conditie. Door het matchen van proefpersonen op aanleg zoals dat in het laatste experiment gebeurde verdwenen de verschillen tussen de instructiecondities namelijk. Het verschil tussen lage aanleg en hoge aanleg was echter binnen beide instructiecondities significant.

In het laatste hoofdstuk worden de belangrijkste bevindingen uit de experimenten naar voren gehaald om verbanden met gerelateerde literatuur te leggen.

De relatie tussen instructie en het leerproces blijkt complexer dan verwacht. Het ontbreken van een één op één relatie tussen geconstateerde fouten en instructie kan verklaard worden vanuit de concepten scaffolding en cognitieve overbelasting. De beide verklaringen sluiten elkaar overigens niet uit.

Feedback wordt ook in de literatuur gezien als een belangrijk element voor het afstemmen van met name de reproductieve aspecten van vaardigheden. Een interessante vraag die hierbij opkomt is of taken waar zowel productieve en reproductieve aspecten aan onderscheiden kunnen worden mogelijk optimaal ondersteund kunnen worden met behulp van 'blended learning'.

Het is daarbij van belang te realiseren dat eigenschappen van de leerling (zoals aanleg) een grote rol kunnen spelen. Wat dat betreft is er een grote behoefte aan een alomvattend kader waarbinnen training voor high-performance taken ontwikkeld kan worden.

Appendix A

Detailed technical specifications of the Low Cost Simulator (LOCS)

Item	Type
Display (5x)	Sony GDM W900 24"
Image generator PC (1x)	TDZ 2000 Pentium II 2x300 Mhz, 128 Mb, 4.3 Gb, Intergraph VX25 64 Mb, Windows NT 4.0
Image generator PC (2x)	GCC Pentium II 450 Mhz, 128 Mb, 4.3 Gb, Accelgraphics 52 Mb, Windows NT 4.0
Model/supervisor PC (2x)	GCC Pentium II 450 Mhz, 64 Mb, 3.2 Gb, Iiyama 17' monitor, DOS 6.22.
Sound/dashboard PC (1x)	GCC Pentium II 350 Mhz, 64 Mb, 3.2 Gb, 32 x CD-ROM, Soundblaster LIVE! PCI, Iiyama 17' monitor, WinNT 4.0
AD converter	DT 2801/5716a
DAC	DT 2816
Headtracker*	Puppetworks Mechanical Headtracking system
Mock-up	Steering wheel, pedals, car seat, gear stick, speedometer
Powersteer	Actuator
Seatshaker*	Linear actuators w. bearings (x,y movement)
Mock up frame	Aluminum
Sound	Amplifier + 4 speakers
Software	Multigen Vega

* not used in the present experiment

Appendix B

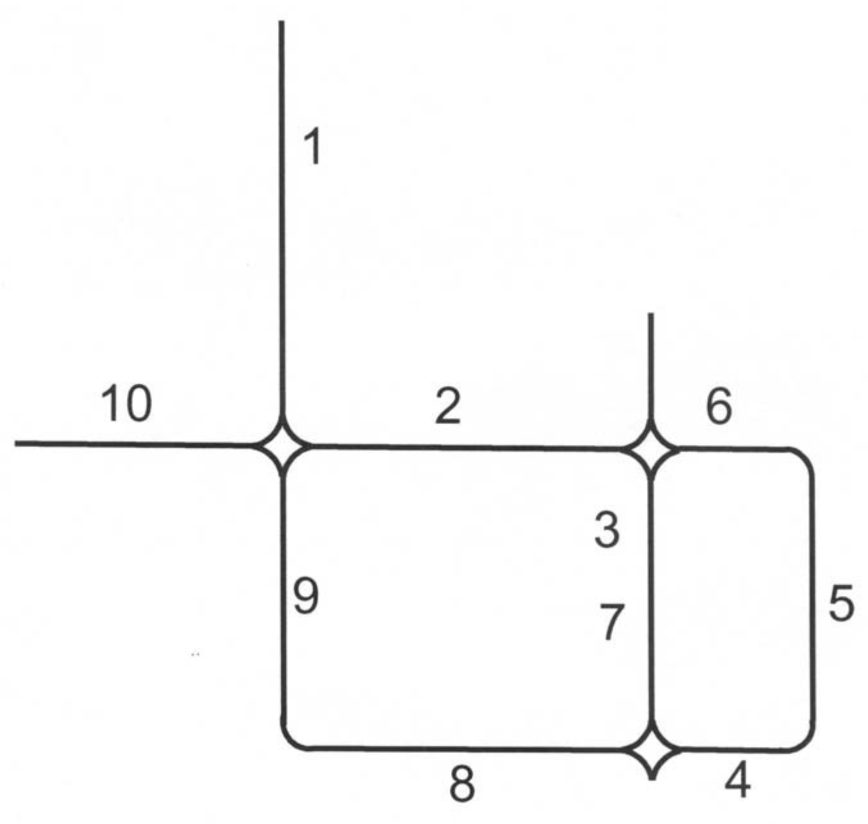
Description of a scenario (example)

The following description has been derived from phase 2b. In this phase subjects had to drive mainly on 50 km/h roads. Only first and second gears were used. There was other traffic. Driving the scenario took about 7 minutes. The numbers in the second column of this description refer to the numbers in the figure on the next page.

Starting point	X	Car placed on the right shoulder, gear neutral. Scenario starts after a recording tells the subject "turn to your lane".
Straight driving for 400m	1	A car drives in front of the simulator car (SC) A car drives in the opposite lane
Left turn (intersection)		A car from the opposite direction crosses the intersection
Straight driving for 400m	2	A car drives in the opposite lane
Right turn (intersection)		A car with right of way comes from the left at 80 km/h
Straight driving for 350m	3	A car drives in the opposite lane
Left turn (intersection)		A car comes from the opposite direction at 80 km/h (forcing the SC to wait) A car comes from left and turns right after the first car passed Another car comes from left and turns left
Straight driving for 180m	4	
Left turn (curve)		A car drives in the opposite lane
Straight driving for 350m	5	A car drives in the opposite lane and wants to overtake a car that has stopped on the road
Left turn (curve)		
Straight driving for 180m	6	
Left turn (intersection)		A car comes from left and turns left at the intersection
Straight driving for 350m	7	A car drives in the opposite lane
Right turn (intersection)		A car from the left waits for SC to turn
Straight driving for 400m	8	
Right turn (curve)		A car drives in the opposite lane
Straight driving for 350m	9	
Left turn (intersection)		A car comes from the left
Straight driving for 300m	10	
End		

The figure below is a detail of the database (see Figure 6). The numbers in the figure match with the second column in the table on the previous page.

(Note that the drawing is only on scale by approximation.)



Appendix C

Questionnaire, (Translated from Dutch)

Date [ ]	Subject No. [ ]	
What age are you?		[…] years
Are you licensed to drive a light motorcycle? (< 50cc) [Dutch: brommer]		Yes [ ] No [ ]
Do you own a light motorcycle?		Yes [ ] No [ ]
If yes, does it require manual gear shifting?		Yes, 2 gears [ ] Yes, more than 2 [ ] No [ ]
If no, did you ever own a light motorcycle that required manual gear shifting?		Yes [ ] No [ ]
What distance do you travel each day? (km)	By motorcycle	None [ ] Less than 10 [ ] More [ ]
	By bicycle	None [ ] Less than 10 [ ] More [ ]
Do you want to take driving lessons?		Yes [ ] Don't know [ ] No [ ]
If yes, on what term		Within 6 months [ ] Within 1 year [ ] Later [ ]
If yes, do you already have a certificate for the theoretical part?		Yes [ ] No [ ]
Do you play computer games?		Never [ ] Less than 1 hour a day [ ] More [ ]
Which of the following input devices do you use on a computer?	Mouse	Mostly not [ ] Now and then [ ] Mostly [ ]
	Joystick	Mostly not [ ] Now and then [ ] Mostly [ ]
	Steering wheel	Mostly not [ ] Now and then [ ] Mostly [ ]

Do you know the following signs?
Please name them: [tested verbally]

1. 2. 3. 4. 5. 6. 7. 8.

Appendix D

Summary of statistical data from discriminant function analyses (exp 1)

Phase 1a*: Straight roads (30 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .65812 approx. F (6,81)=7.0130 p< .0000

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*55.0*	*22*	*18*	*40*
Passed	*91.7*	*4*	*44*	*48*
Total predicted	*75.0*	*26*	*62*	*88*

Phase 1a: *Right curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 1a: *Left curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 1b: *Straight roads (30 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .76645 approx. F (6,89)=4.5200 p< .0005

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*70.83*	*17*	*7*	*24*
Passed	*77.78*	*16*	*56*	*72*
Total predicted	*76.04*	*33*	*63*	*96*

Phase 1b: *Right curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .59170 approx. F (6,29)=3.3353 p< .0127

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*77.78*	*7*	*2*	*9*
Passed	*92.59*	*2*	*25*	*27*
Total predicted	*88.89*	*9*	*27*	*36*

Phase 1b: *Left curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .68044 approx. F (6,41)=3.2092 p< .0112

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*75*	*9*	*3*	*12*
Passed	*80.56*	*7*	*29*	*36*
Total predicted	*79.17*	*16*	*32*	*48*

Phase 2a: *Straight roads (50 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .82201 approx. F (6,89)=3.2118 p< .0067

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*60*	*24*	*16*	*40*
Passed	*78.57*	*12*	*44*	*56*
Total predicted	*70.83*	*36*	*60*	*96*

Phase 2a: *Right curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .59407 approx. F (6,41)=4.6693 p< .0011

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*90*	*18*	*2*	*20*
Passed	*78.57*	*6*	*22*	*28*
Total predicted	*83.33*	*24*	*24*	*48*

Phase 2a: *Left curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .57638 approx. F (6,29)=3.5523 p< .0093

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*66.67*	*10*	*5*	*15*
Passed	*85.71*	*3*	*18*	*21*
Total predicted	*77.78*	*13*	*23*	*36*

Phase 2b: *Straight roads (50 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .77174 approx. F (6,89)=4.3873 p< .0006

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*66.67*	*16*	*8*	*24*
Passed	*76.39*	*17*	*55*	*62*
Total predicted	*73.96*	*33*	*63*	*96*

Phase 2b: *Right curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 2b: *Left curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .51089 approx. F (6,29)=4.6272 p< .0021

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*88.89*	*8*	*1*	*9*
Passed	*88.89*	*3*	*24*	*27*
Total predicted	*88.89*	*11*	*25*	*36*

Phase 3a: *Straight roads (80 km/h)*
Discriminant Function Analysis Summary
Analysis not significant

Phase 3a: *Right curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .33483 approx. F (6,23)=7.6153 p< .0001

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*88.89*	*8*	*1*	*9*
Passed	*90.48*	*2*	*19*	*21*
Total predicted	*90*	*10*	*20*	*30*

Phase 3a: *Left curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 3b: *Straight roads (80 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .73928 approx. F (6,113)=6.6421 p< .0000

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*61.11*	*22*	*14*	*36*
Passed	*79.76*	*17*	*67*	*84*
Total predicted	*74.17*	*39*	*81*	*120*

Phase 3b: *Right curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 3b: *Left curves*
Discriminant Function Analysis Summary
Analysis not significant

Phase 4a: *Straight roads (80 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .44971 approx. F (6,17)=3.4670 p< .0201

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*66.67*	*2*	*1*	*3*
Passed	*100*	*0*	*21*	*21*
Total predicted	*95.83*	*2*	*22*	*24*

Phase 4a: *Straight roads (50 km/h)*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .07123 approx. F (6,41)=89.099 p< .0000

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed				
Passed				
Total predicted				

Phase 4a: *Right curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .10252 approx. F (6,25)=36.476 p< .0000

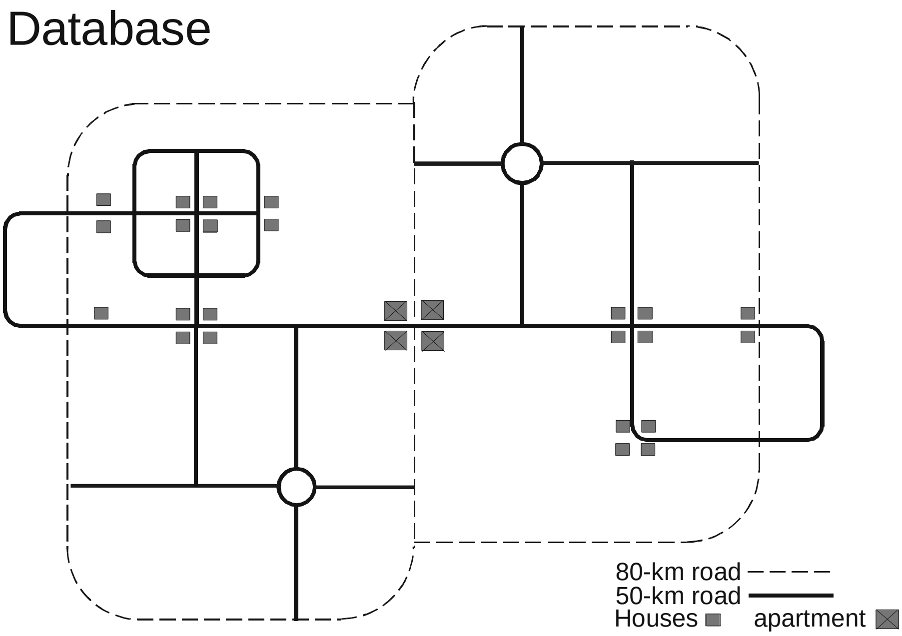
Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*100*	*4*	*0*	*4*
Passed	*100*	*0*	*28*	*28*
Total predicted	*100*	*4*	*28*	*32*

Phase 4a: *Left curves*
Discriminant Function Analysis Summary
No. of variables in model: 6; Grouping: 2 groups
Wilks' Lambda: .06979 approx. F (6,17)=37.766 p< .0000

Classification matrix	*Percent Correct*	*Failed p= 0.5*	*Passed p= 0.5*	*Total observed*
Failed	*100*	*3*	*0*	*3*
Passed	*100*	*0*	*21*	*21*
Total predicted	*100*	*3*	*21*	*24*

Phase 4b:
Discriminant Function Analysis Summary
None of the analyses were significant

Appendix E



The revised LOCS database plan as it was used in experiment 2 and 3. Compared with the old plan changes were applied to the roundabouts and the number of buildings. See Figure 6 and section 6.2.1.2 (Database).

Curriculum Vitae

Martijn van Emmerik werd op 2 juni 1972 geboren in Haarlem. Hij volgde zijn VWO opleiding aan de scholengemeenschap ‘Het Rijnlands Sassenheim’ in Sassenheim. In 1990 begon hij met een studie sociologie aan de Universiteit Utrecht maar stapte daar halverwege het jaar over naar psychologie. Deze studie rondde hij in 1996 af in twee afstudeerrichtingen (Cognitieve Ergonomie en Experimentele Functieleer). Voor het empirische deel van zijn studie werkte hij negen maanden op de afdeling Vaardigheden van TNO Technische Menskunde in Soesterberg. Na korte tijd als uitzendkracht op dit instituut gewerkt te hebben trad hij in 1997 in dienst van de Universiteit Twente waar hij als AIO voor de faculteit Toegepaste Onderwijskunde (vakgroep Instructie Technologie) werkte. Zijn werkzaamheden voerde hij uit bij de afdeling Training en Opleiding van TNO-TM. Aansluitend hieraan kwam hij in 2001 in dienst bij TNO-TM en werkte onder meer aan projecten op het gebied van virtuele (automatische) instructie in simulatoren; verificatie, validatie en accreditatie; evaluatie van trainingsfaciliteiten en opleidingen; en sport.